NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME

RELATED APPLICATIONS

This application claims priority from Provisional Applications U.S.S.N. 60/185,674, filed February 28, 2000; U.S.S.N. 60/185,535, filed February 28, 2000; U.S.S.N. 60/186,717, filed March 3, 2000; U.S.S.N. 60/186,585, filed March 3, 2000; U.S.S.N. 60/186,604, filed March 3, 2000; U.S.S.N. 60/218,323, filed July 14, 2000; U.S.S.N. 60/220,517, filed July 24, 2000; U.S.S.N. 60/186,584, filed March 3, 2000; U.S.S.N. 60/186,827, filed March 3, 2000; U.S.S.N. 60/260,020, filed on January 5, 2001; U.S.S.N. 60/186,716, filed on March 3, 2000; U.S.S.N. 60/218,435, filed July 14, 2000; U.S.S.N. 60/186,715, filed March 3, 2000; U.S.S.N. 60/223,897, filed August 9, 2000; U.S.S.N. 60/264,849, filed January 26, 2001; U.S.S.N. 60/186,719, filed March 3, 2000, U.S.S.N. 60/259,031, filed December 28, 2000, USSN 60/215,855 filed July 3, 2000, USSN 09/795,271 filed February 27, 2001, and USSN 09/809,476 filed March 15, 2001, each of which is incorporated by reference in its entirety.

BACKGROUND OF THE INVENTION

The invention generally relates to nucleic acids and polypeptides. More particularly, the invention relates to nucleic acids encoding novel G-protein coupled receptor (GPCR) polypeptides, as well as vectors, host cells, antibodies, and recombinant methods for producing these nucleic acids and polypeptides.

SUMMARY OF THE INVENTION

The invention is based in part upon the discovery of nucleic acid sequences encoding novel polypeptides. The novel nucleic acids and polypeptides are referred to herein as GPCRX, or GPCR1, GPCR2, GPCR3, GPCR4, GPCR5, GPCR6, GPCR7, GPCR8, GPCR9, and GPCR10 nucleic acids and polypeptides. These nucleic acids and polypeptides, as well as derivatives, homologs, analogs and fragments thereof, will hereinafter be collectively designated as "GPCRX" nucleic acid or polypeptide sequences.

In one aspect, the invention provides an isolated GPCRX nucleic acid molecule encoding a GPCRX polypeptide that includes a nucleic acid sequence that has identity to the nucleic acids disclosed in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84. In some embodiments, the GPCRX nucleic acid molecule will hybridize under stringent

conditions to a nucleic acid sequence complementary to a nucleic acid molecule that includes a protein-coding sequence of a GPCRX nucleic acid sequence. The invention also includes an isolated nucleic acid that encodes a GPCRX polypeptide, or a fragment, homolog, analog or derivative thereof. For example, the nucleic acid can encode a polypeptide at least 80% identical to a polypeptide comprising the amino acid sequences of SEQ ID NOS:2, 4, 6, 8, 10, 12, 17, 19, 21, 23, 25, 29, 31, 33, 35, 37, 83, and 85. The nucleic acid can be, for example, a genomic DNA fragment or a cDNA molecule that includes the nucleic acid sequence of any of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84.

Also included in the invention is an oligonucleotide, *e.g.*, an oligonucleotide which includes at least 6 contiguous nucleotides of a GPCRX nucleic acid (*e.g.*, SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84) or a complement of said oligonucleotide.

Also included in the invention are substantially purified GPCRX polypeptides (SEQ ID NOS:2, 4, 6, 8, 10, 12, 17, 19, 21, 23, 25, 29, 31, 33, 35, 37, 83, and 85). In certain embodiments, the GPCRX polypeptides include an amino acid sequence that is substantially identical to the amino acid sequence of a human GPCRX polypeptide.

The invention also features antibodies that immunoselectively bind to GPCRX polypeptides, or fragments, homologs, analogs or derivatives thereof.

In another aspect, the invention includes pharmaceutical compositions that include therapeutically- or prophylactically-effective amounts of a therapeutic and a pharmaceutically-acceptable carrier. The therapeutic can be, *e.g.*, a GPCRX nucleic acid, a GPCRX polypeptide, or an antibody specific for a GPCRX polypeptide. In a further aspect, the invention includes, in one or more containers, a therapeutically- or prophylactically-effective amount of this pharmaceutical composition.

In a further aspect, the invention includes a method of producing a polypeptide by culturing a cell that includes a GPCRX nucleic acid, under conditions allowing for expression of the GPCRX polypeptide encoded by the DNA. If desired, the GPCRX polypeptide can then be recovered.

In another aspect, the invention includes a method of detecting the presence of a GPCRX polypeptide in a sample. In the method, a sample is contacted with a compound that selectively binds to the polypeptide under conditions allowing for formation of a complex between the polypeptide and the compound. The complex is detected, if present, thereby identifying the GPCRX polypeptide within the sample.

The invention also includes methods to identify specific cell or tissue types based on their expression of a GPCRX.

Also included in the invention is a method of detecting the presence of a GPCRX nucleic acid molecule in a sample by contacting the sample with a GPCRX nucleic acid probe or primer, and detecting whether the nucleic acid probe or primer bound to a GPCRX nucleic acid molecule in the sample.

In a further aspect, the invention provides a method for modulating the activity of a GPCRX polypeptide by contacting a cell sample that includes the GPCRX polypeptide with a compound that binds to the GPCRX polypeptide in an amount sufficient to modulate the activity of said polypeptide. The compound can be, *e.g.*, a small molecule, such as a nucleic acid, peptide, polypeptide, peptidomimetic, carbohydrate, lipid or other organic (carbon containing) or inorganic molecule, as further described herein.

Also within the scope of the invention is the use of a therapeutic in the manufacture of a medicament for treating or preventing disorders or syndromes including, *e.g.*, diabetes, metabolic disturbances associated with obesity, the metabolic syndrome X, anorexia, wasting disorders associated with chronic diseases, metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders, or other disorders related to cell signal processing and metabolic pathway modulation. The therapeutic can be, *e.g.*, a GPCRX nucleic acid, a GPCRX polypeptide, or a GPCRX-specific antibody, or biologically-active derivatives or fragments thereof.

For example, the compositions of the present invention will have efficacy for treatment of patients suffering from: developmental diseases, MHCII and III diseases (immune diseases), taste and scent detectability Disorders, Burkitt's lymphoma, corticoneurogenic disease, signal transduction pathway disorders, Retinal diseases including those involving photoreception, Cell growth rate disorders; cell shape disorders, feeding disorders; control of feeding; potential obesity due to over-eating; potential disorders due to starvation (lack of appetite), noninsulin-dependent diabetes mellitus (NIDDM1), bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease; multiple sclerosis; Albright Hereditary Ostoeodystrophy, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy, and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression,

delirium, dementia, severe mental retardation. Dentatorubro-pallidoluysian atrophy (DRPLA) Hypophosphatemic rickets, autosomal dominant (2) Acrocallosal syndrome and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome and/or other pathologies and disorders of the like.

The polypeptides can be used as immunogens to produce antibodies specific for the invention, and as vaccines. They can also be used to screen for potential agonist and antagonist compounds. For example, a cDNA encoding GPCRX may be useful in gene therapy, and GPCRX may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to Neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease; multiple sclerosis; and Treatment of Albright Hereditary Ostoeodystrophy, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy, and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome and/or other pathologies and disorders.

The invention further includes a method for screening for a modulator of disorders or syndromes including, *e.g.*, diabetes, metabolic disturbances associated with obesity, the metabolic syndrome X, anorexia, wasting disorders associated with chronic diseases, metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders or other disorders related to cell signal processing and metabolic pathway modulation. The method includes contacting a test compound with a GPCRX polypeptide and determining if the test compound binds to said GPCRX polypeptide. Binding of the test compound to the GPCRX polypeptide indicates the test compound is a modulator of activity, or of latency or predisposition to the aforementioned disorders or syndromes.

Also within the scope of the invention is a method for screening for a modulator of activity, or of latency or predisposition to an disorders or syndromes including, *e.g.*, diabetes, metabolic disturbances associated with obesity, the metabolic syndrome X, anorexia, wasting disorders associated with chronic diseases, metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders or other disorders

related to cell signal processing and metabolic pathway modulation by administering a test compound to a test animal at increased risk for the aforementioned disorders or syndromes. The test animal expresses a recombinant polypeptide encoded by a GPCRX nucleic acid. Expression or activity of GPCRX polypeptide is then measured in the test animal, as is expression or activity of the protein in a control animal which recombinantly-expresses GPCRX polypeptide and is not at increased risk for the disorder or syndrome. Next, the expression of GPCRX polypeptide in both the test animal and the control animal is compared. A change in the activity of GPCRX polypeptide in the test animal relative to the control animal indicates the test compound is a modulator of latency of the disorder or syndrome.

In yet another aspect, the invention includes a method for determining the presence of or predisposition to a disease associated with altered levels of a GPCRX polypeptide, a GPCRX nucleic acid, or both, in a subject (*e.g.*, a human subject). The method includes measuring the amount of the GPCRX polypeptide in a test sample from the subject and comparing the amount of the polypeptide in the test sample to the amount of the GPCRX polypeptide present in a control sample. An alteration in the level of the GPCRX polypeptide in the test sample as compared to the control sample indicates the presence of or predisposition to a disease in the subject. Preferably, the predisposition includes, *e.g.*, diabetes, metabolic disturbances associated with obesity, the metabolic syndrome X, anorexia, wasting disorders associated with chronic diseases, metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders. Also, the expression levels of the new polypeptides of the invention can be used in a method to screen for various cancers as well as to determine the stage of cancers.

In a further aspect, the invention includes a method of treating or preventing a pathological condition associated with a disorder in a mammal by administering to the subject a GPCRX polypeptide, a GPCRX nucleic acid, or a GPCRX-specific antibody to a subject (*e.g.*, a human subject), in an amount sufficient to alleviate or prevent the pathological condition. In preferred embodiments, the disorder, includes, *e.g.*, diabetes, metabolic disturbances associated with obesity, the metabolic syndrome X, anorexia, wasting disorders associated with chronic diseases, metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders.

In yet another aspect, the invention can be used in a method to identity the cellular receptors and downstream effectors of the invention by any one of a number of techniques

commonly employed in the art. These include but are not limited to the two-hybrid system, affinity purification, co-precipitation with antibodies or other specific-interacting molecules.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

DETAILED DESCRIPTION OF THE INVENTION

The invention is based, in part, upon the discovery of novel nucleic acid sequences that encode novel polypeptides. The novel nucleic acids and their encoded polypeptides are referred to individually as GPCR1, GPCR2, GPCR3, GPCR4, GPCR5, GPCR6, GPCR7, GPCR8, GPCR9, and GPCR10. The nucleic acids, and their encoded polypeptides, are collectively designated herein as "GPCRX".

The novel GPCRX nucleic acids of the invention include the nucleic acids whose sequences are provided in Tables 1A, 1E, 1I, 2A, 2I, 3A, 4A, 4C, 4G, 5A, 5C, 5G, 6A, 7A, 8A, 9A, 10A, 10C and 10F, inclusive ("Tables 1A - 10F"), or a fragment, derivative, analog or homolog thereof. The novel GPCRX proteins of the invention include the protein fragments whose sequences are provided in Tables 1B, 1F, 1J, 2B, 2J, 3B, 4B, 4H, 5B, 5D, 5H, 6B, 7B, 8B, 9B, 10B, 10D, and 10G, inclusive ("Tables 1B - 10G"). The individual GPCRX nucleic acids and proteins are described below. Within the scope of this invention is a method of using these nucleic acids and peptides in the treatment or prevention of a disorder related to cell signaling or metabolic pathway modulation.

G-Protein Coupled Receptor proteins (GPCRs) have been identified as a large family of G protein-coupled receptors in a number of species. These receptors share a seven transmembrane domain structure with many neurotransmitter and hormone receptors, and are likely to underlie the recognition and G-protein-mediated transduction of various signals. Human GPCR generally do not contain introns and belong to four different gene subfamilies, displaying great sequence variability. These genes are dominantly expressed in olfactory

epithelium. See, *e.g.*, Ben-Arie et al., Hum. Mol. Genet. 1994 3:229-235; and, Online Mendelian Inheritance in Man (OMIM) entry # 164342 (http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?).

The olfactory receptor (OR) gene family constitutes one of the largest GPCR multigene families and is distributed among many chromosomal sites in the human genome. See Rouquier et al., Hum. Mol. Genet. 7(9):1337-45 (1998); Malnic et al., Cell 96:713-23 (1999). Olfactory receptors constitute the largest family among G protein-coupled receptors, with up to 1000 members expected. See Vanderhaeghen et al., Genomics 39(3):239-46 (1997); Xie et al., Mamm. Genome 11(12):1070-78 (2000); Issel-Tarver et al., Proc. Natl. Acad. Sci. USA 93(20):10897-902 (1996). The recognition of odorants by olfactory receptors is the first stage in odor discrimination. See Krautwurst et al., Cell 95(7):917-26 (1998); Buck et al., Cell 65(1):175-87 (1991). Many ORs share some characteristic sequence motifs and have a central variable region corresponding to a putative ligand binding site. See Issel-Tarver et al., Proc. Natl. Acad. Sci. USA 93:10897-902 (1996).

Other examples of seven membrane spanning proteins that are related to GPCRs are chemoreceptors. See Thomas et al., Gene 178(1-2):1-5 (1996). Chemoreceptors have been identified in taste, olfactory, and male reproductive tissues. See *id.*; Walensky et al., J. Biol. Chem. 273(16):9378-87 (1998); Parmentier et al., Nature 355(6359):453-55 (1992); Asai et al., Biochem. Biophys. Res. Commun. 221(2):240-47 (1996).

GPCR1

GPCR1 includes a family of three novel G-protein coupled receptor ("GPCR") proteins disclosed below. The disclosed proteins have been named GPCR1a, GPCR1b, and GPCR1c and are related to olfactory receptors.

GPCR1a

A disclosed GPCR1a nucleic acid of 1019 nucleotides is shown in Table 1A. The disclosed GPCR1a open reading frame ("ORF") begins at the ATG initiation codon at nucleotides 27-29, shown in bold in Table 1A. The encoded polypeptide is alternatively referred to herein as GPCR1a or as ba113a10_B. The disclosed GPCR1a ORF terminates at a TAG codon at nucleotides 984-986. As shown in Table 1A, putative untranslated regions 5' to the start codon and 3' to the stop codon are underlined, and the start and stop codons are in bold letters.

Table 1A. GPCR1a nucleotide sequence (SEQ ID NO:1).

CCTTCAGTTGACAGAGGAGATACACT**ATG**GTAAGTGCCAATCAGACAGCCTCTGTGACCGAGTTTATTC
TCCTGGGCCTCTCTGCCCACCCAAAGCTGGAGAAAACGTTCTTTGTGCTCATCCTGCTGATGTACCTGG
TGATCCTACTGGGCAATGGGGTCCTCATCCTGATGACTGTGTCCAACTCCCACCTGCACATGCCCATGT
ACTTCTTCCTGGGGAACCTCTCCTTCCTGGACATCTGCTATACAACATCCTCAGTCCCCCTCATCCTTG
ACAGCTTCTTGACCCCCAGGAAAACCATCTCCTTCTCAGCCTGTGCAGTGCAGATGTTCCTCTCCTTTG
CCATGGGAGCCACAGAGTGTGTTCTCCTGAGCATGATGGCGTTTGATCGCTACGTGGCCATCTGCAACC
CCCTTAGGTACCCTGTGGTCATGAGCAAGGCTGCCTACATGCCCCATAAGGCTGCCGGCTCCTGGGTAG
CTGGAAGCACTGCTTCCATGGTGCAGACATCCCTTGCAATGAGGCTGCCCTTCTGTGGAGACAACATCA
TCAACCACTTCACCTGTGAGATTCTGGCTGTCCTGAAGTTGGCCTGTGCTGATATCTCTGTCAATGTGA
TCAGTATGGGAGTGACCAATGTGATCTTCCTGGGGGTCCCGGTTCTGTTCATCTCTTTCTCCTATGTCT
TCATCATTGCCACCATCCTGAGGATCCCCTCAGCTGAGGGGAGGAAAAAGGCCTTCTCCACCTGCTCTG
CCCACCTCACAGTCGTGGTCATCTTCTATGGGACCATCCTCTTCATGTATGGGAAGCCCAAGTCTAAGG
ACCCGCTGGGGGCAGACAAGCAAGACCTTGCAGACAAACTCATTTCCCTTTTCTATGGGGTGGTGACCC
CCATGCTCAACCCCATCATCTACAGCCTGAGGAACAAGGATGTAAAGGCTGCTGTGAGGGACTTGATAT
TTCAGAAATGCTTTGCC**TAG**TGATGTTTGGGGGAACAGATGTCCTCATAGCTC

A disclosed encoded GPCR1a protein has 319 amino acid residues, referred to as the GPCR1a protein. The GPCR1a protein was analyzed for signal peptide prediction and cellular localization. SignalP results predict that GPCR1a is cleaved between position 44 and 55 of SEQ ID NO:2, *i.e.*, at the slash in the amino acid sequence GNG-VL. Psort and Hydropathy profiles also predict that GPCR1 contains a signal peptide and is likely to be localized at the plasma membrane (certainty of 0.6000). The disclosed GPCR1 polypeptide sequence is presented in Table 1B using the one-letter amino acid code.

Table 1B. Encoded GPCR1a protein sequence (SEQ ID NO:2).

MVSANQTASVTEFILLGLSAHPKLEKTFFVLILLMYLVILLGNG/VLILMTVSNSHLHMPMYFFLGNLS
FLDICYTTSSVPLILDSFLTPRKTISFSACAVQMFLSFAMGATECVLLSMMAFDRYVAICNPLRYPVVM
SKAAYMPHKAAGSWVAGSTASMVQTSLAMRLPFCGDNIINHFTCEILAVLKLACADISVNVISMGVTNV
IFLGVPVLFISFSYVFIIATILRIPSAEGRKKAFSTCSAHLTVVVIFYGTILFMYGKPKSKDPLGADKQ
DLADKLISLFYGVVTPMLNPIIYSLRNKDVKAAVRDLIFQKCFA

GPCR1a was initially identified on chromosome 9 with a TblastN analysis of a proprietary sequence file for a G-protein coupled receptor probe or homolog, which was run against the Genomic Daily Files made available by GenBank. A proprietary software program (GenScan™) was used to further predict the nucleic acid sequence and the selection of exons. The resulting sequences were further modified by means of similarities using BLAST searches. The sequences were then manually corrected for apparent inconsistencies, thereby obtaining the sequences encoding the full-length protein.

A region of the GPCR1a nucleic acid sequence has 873 of 1016 bases (85%) identical to a sequence coding for a partial *Mus musculus* olfactory receptor mRNA (1731 bp), with an E-value of $3.6e^{-160}$ (GENBANK-ID: Aj133427). In all BLAST alignments herein, the "E-value" or "Expect" value is a numeric indication of the probability that the aligned sequences could have

achieved their similarity to the BLAST query sequence by chance alone, within the database that was searched. For example, the probability that the subject ("Sbjct") retrieved from the GPCR1a BLAST analysis, *e.g.*, the *Mus musculus* olfactory receptor, matched the Query GPCR1a sequence purely by chance is $3.6x10^{-160}$.

A BLASTX search was performed against public protein databases. The full amino acid sequence of the protein of the invention was found to have 274 of 316 amino acid residues (86%) identical to, and 295 of 316 residues (93%) positive with, the 319 amino acid olfactory receptor protein from *Mus musculus* (ptnr:TREMBLNEW-ACC:CAB55592, E= 1.8 e-142). The disclosed GPCR1a protein (SEQ ID NO:2) has good identity with a number of olfactory receptor proteins. For example, GPCR1a has 256/316 (81%) amino acids identical with the 319 amino *acid Mus musculus* olfactory receptor 37a protein, and 376/316 (87%) amino acids identical, to (Expect = e-128, gi|11276075|ref|NP_062346.1|) olfactory receptor 37a from *Mus musculus*. The disclosed protein is also similar to the olfactory proteins disclosed in Table 1C.

Table 1C. BLAST results for GPCR1a

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
Gi|11464981|ref|NP_062349.1|	Olfactory Receptor (OR) 37e *Mus musculus*	319	255/318 (80%)	275/318 (86%)	e-127
Gi|11276077|ref|NP_062347.1|	OR 37b *Mus musculus*	318	248/317 (78%)	276/317 (86%)	e-127
Gi|11276079|ref|NP_062348.1|	OR 37c *Mus musculus*	318	251/314 (79%)	274/314 (86%)	e-127
Gi|10092669|ref|NP_063950.1|	OR Family 2, Subfamily S, member 2 *Homo sapiens*	309	238/306 (77%)	261/306 (84%)	e-118
Gi|3769624|gb|AAC64588.1| (AF091565)	OR *Rattus norvegicus*	227	206/227 (90%)	217/227 (94%)	e-102

Other BLAST results include sequences from the Patp database, which is a proprietary database that contains sequences published in patents and patent publications. Patp results include those listed in Table 1D.

Table 1D. Patp alignments of GPCR1a

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob. P(N)
patp:B43266 Human ORFX ORF3030 polypeptide sequence SE.	+3	697	6.2e-68
patp:R27872 Odorant receptor clone I7 - Rattus rattus,..	+3	693	1.6e-67
patp:Y83389 Olfactory receptor protein OLF-4 - H. sapiens.	+3	677	8.1e-66
patp:Y83387 Olfactory receptor protein OLF-2 - H. sapiens.	+3	674	1.7e-65
patp:R27868 Odorant receptor clone F5 - Rattus rattus,..	+3	664	1.9e-64
patp:Y83390 Olfactory receptor protein OLF-5 - H. sapiens.	+3	660	5.2e-64

For example, a BLAST against patp: R27872, a 327 amino acid odorant receptor clone from *Rattus rattus*, produced 140/310 (45%) identity, and 197/310 (63%) positives ($E = 1.6e^{-67}$). PCT publication WO 92/17585. See also WO 00/58473.

GPCR1b

A disclosed GPCR1b (also referred to as ba32713_A) nucleic acid of 1015 nucleotides is shown in Table 1E. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 17-19 and ending with a TAG codon at nucleotides 971-973. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in Table 1E, and the start and stop codons are in bold letters.

Table 1E. GPCR1b Nucleic acid sequence (SEQ ID NO:3).

ACAGAGGAGATACACT**ATG**GTAAGTGCCAATCAGACAGCCTCTGTGACCGAGTTTATTCTCCTGGGCCTC
TCTGCCCACCCAAAGCTGGAGAAAACGTTCTTTGTGCTCATCCTGCTGATGTACCTGGTGATCCTACTGG
GCAATGGGGTCCTCATCCTGATGACTGTGTCCAACTCCCACCTGCACATGCCCATGTACTTCTTCCTGGG
GAACCTCTCCTTCCTGGACATCTGCTATACAACATCCTCAGTCCCCCTCATCCTTGACAGCTTCTTGACC
CCCAGGAAAACCATCTCCTTCTCAGCCTGTGCAGTGCAGATGTTCCTCTCCTTTGCCATGGGAGCCACAG
AGTGTGTTCTCCTGAGCATGATGGCGTTTGATCGCTACGTGGCCATCTGCAACCCCCTTAGGTACCCTGT
GGTCATGAGCAAGGCTGCCTACATGCCCATAGCTGCCGGCTCCTGGGTAGCTGGAAGCACTGCTTCCATG
GTGCAGACATCCCTTGCAATGAGGCTGCCCTTCTGTGGAGACAACATCATCAACCACTTCACCTGTGAGA
TTCTGGCTGTCCTGAAGTTGGCCTGTGCTGATATCTCTGTCAATGTGATCAGTATGGGAGTGACCAATGT
GATCTTCCTGGGGGTCCCGGTTCTGTTCATCTCTTTCTCCTATGTCTTCATCATTGCCACCATCCTGAGG
ATCCCCTCAGCTGAGGGGAGGAAAAAGGCCTTCTCCACCTGCTCTGCCCACCTCACAGTCGTGGTCATCT
TCTATGGGACCATCCTCTTCATGTATGGGAAGCCCAAGTCTAAGGACCCGCTGGGGGCAGACAAGCAAGA
CCTTGCAGACAAACTCATTTCCCTTTTCTATGGGGTGGTGACCCCCATGCTCAACCCCATCATCTACAGC
CTGAGGAACAAGGATGTAAAGGCTGCTGTGAGGGACTTGATATTTCAGAAATGCTTTGCCTAGTGATGTT
TGGGGGAACAGATGTCCTCA**TAG**CTCTTTGCCTCT

In a search of sequence databases, it was found, for example, that the nucleic acid sequence has 869 of 1015 bases (85%) identical to a 1731 bp *Mus musculus* OR 37d pseudogene (GENBANK-ID: MMU133427|acc:AJ133427, E = 5.5 e-161). It was also found that the nucleic acid has 505 of 791 bases identical (63%) to a partial human mRNA for olfactory receptor protein (GENBANK-ID:HSOLFMF|acc:Y14442, E= 9.5 e-49).

The encoded protein having 318 amino acid residues is presented using the one-letter code in Table 1F. The full amino acid sequence of the protein of the invention was found to have 274 of 315 amino acid residues (86%) identical to, and 296 of 315 residues (93%) positive with, a 319 amino acid residue olfactory receptor protein from *Mus musculus* (ptnr: TREMBLNEW-ACC:CAB55592, E= 6.3 e-144.

The disclosed GPCR1b protein differs from the disclosed GPCR1a protein at only two positions. At positions 145 and 146, GPCR1a has HK, while GPCR1b has a deletion (Δ) and an I.

Table 1F. Encoded GPCR1b protein sequence (SEQ ID NO:4).
MVSANQTASVTEFILLGLSAHPKLEKTFFVLILLMYLVILLGNG/VLILMTVSNSHLHMPMYFFLGNLSFL DICYTTSSVPLILDSFLTPRKTISFSACAVQMFLSFAMGATECVLLSMMAFDRYVAICNPLRYPVVMSKA AYMPIAAGSWVAGSTASMVQTSLAMRLPFCGDNIINHFTCEILAVLKLACADISVNVISMGVTNVIFLGV PVLFISFSYVFIIATILRIPSAEGRKKAFSTCSAHLTVVVIFYGTILFMYGKPKSKDPLGADKQDLADKL ISLFYGVVTPMLNPIIYSLRNKDVKAAVRDLIFQKCFA

A PSORT analysis predicts that the ba32713_A protein (GPCR1b) is localized in the plasma membrane with a certainty of 0.6000, or with lower certainty in the mitochondrial inner membrane, the mitochondrial intermembrane space or the Golgi body. It is also predicted that the protein has a signal peptide with the most likely cleavage site between residues 44 and 45: GNG-VL, indicated by a slash in Table 1F.

A BLASTX search was performed against public protein databases. The full amino acid sequence of the protein of the invention was found to have 274 of 315 amino acid residues (86%) identical to, and 296 of 315 residues (93%) positive with, the 319 amino acid olfactory receptor protein from *Mus musculus* (ptnr:TREMBLNEW-ACC:CAB55592, E= 6.3 e-144). The disclosed GPCR1b protein (SEQ ID NO:4) has good identity with a number of olfactory receptor proteins, as shown in Table 1G.

Table 1G. BLAST results for GPCR1b

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
Gi|11276075|ref|NP_062346.1|	OR 37a *Mus musculus*	319	256/315 (81%)	277/315 (87%)	e-129
Gi|11464981|ref|NP_062349.1|	(OR) 37e *Mus musculus*	319	255/317 (80%)	276/317 (86%)	e-128
Gi|11276077|ref|NP_062347.1|	OR 37b *Mus musculus*	318	248/316 (78%)	277/316 (87%)	e-128
Gi|11276079|ref|NP_062348.1|	OR 37c *Mus musculus*	318	251/313 (80%)	275/313 (87%)	e-128
Gi|10092669|ref|NP_063950.1|	OR Family 2, Subfamily S, member 2 *Homo sapiens*	309	238/305 (78%)	262/305 (85%)	e-119

Patp results include those listed in Table 1H.

Table 1H. Patp alignments of GPCR1b

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob. P(N)
patp:B43266 Human ORFX ORF3030 polypeptide sequence SE.	+2	705	8.8e-69
patp:R27872 Odorant receptor clone I7 - Rattus rattus,..	+2	703	1.4e-68
patp:Y83389 Olfactory receptor protein OLF-4 - H. sapiens.	+2	690	3.4e-67
patp:Y83387 Olfactory receptor protein OLF-2 - H. sapiens.	+2	687	7.1e-67
patp:R27868 Odorant receptor clone F5 - Rattus rattus,..	+2	677	8.1e-66
patp:Y83390 Olfactory receptor protein OLF-5 - H. sapiens.	+2	673	2.2e-65

GPCR1c

A disclosed GPCR1c (also referred to as ba113a10_C) nucleic acid of 1003 nucleotides is shown in Table 1I. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 26-28 and ending with a TGA codon at nucleotides 974-976. Putative untranslated regions 5' to the start codon and 3' to the stop codon are underlined in Table 1I and the start and stop codons are in bold letters.

In a search of sequence databases, it was found, for example, that the nucleic acid sequence has 719 of 899 bases (79%) identical to a *Mus musculus* GPCR mRNA (GENBANK-ID: AJ133428, E = 1.9e-120).

Table 1I. GPCR1c Nucleic acid sequence (SEQ ID NO:5).

TTTGTACAAGTGACATAGAAACACC**ATG**GTCAGTTCCAATCAGACCTCCCCTGTGCTGGGGTTCCTTCTCC
TGGGGCTCTCTGCCCATCCAAAGCTGGAGAAGACATTCTTCGTGCTCATCCTGCTGATGTACCTGGTGATC
CTACTGGGCAATGGGGTCCTCATCCTGGTGACCATCCTTGACTCCCGCCTGGACACACCCATGTACTTCTT
CCTGGGGAACCTCTCCTTCCTGGACATCTGCTATACAACCTCCTCATCCTTGACAGCTTCCCTGACCCCCA
GGAAAACCATCTCCTTCTCAGCCTGTGCAGTACAGATGTTCCTCTCCCTTGCCATGGGAGCCACAGAGTGT
GTTCTCCTGAGCATGATGGCGTTTGATCGCTACGTGGCCATCTGCAACCCCCTTTGGTACCCTGAAGTCAT
GAACAAAGCTACTTATGTGCCCATGGCTGCTGGCTCCTGGGTAGCTGGAAGCCTCACTGCCATGGTGCAGA
CACCCCTTGCATTGAGGCTGCCCTTCTGTGGAGACAACATCATCAATCACTTCACCTGTGAGATTCTGGCT
GTCCTGAAGTTGGCCTGTGCTGATATCTCTGTCAATGTGATCAGTATGGGAGTGGCCAATGTGATCTTCCT
GGGGGTCCCTGTTCTGTTCATCTCTTTCTCCTATGTCTTCATCATTGCCACCATCCTGAGGATCCCCTCAG
CTGAGGGGAGGAAAAAGGCCTTCTCCACCTGCTCTGCCCACCTCACTGTCGTGATCGTCTTCTACGGGACC
ATCCTCTTCATGTACGGGAAGCCCAAGTCTAAGGACCCACTGGGAGCAGACAAACAGGACCTTGCAGACAA
ACTCATTTCCCTTTTCTATGGGGTGGTGACCCCCATGCTCAACCCCATCATCTACAGCCTGAGGAACAAGG
AAGTGAAGGCTGCTGTGAGGAACCTGGTATTTCAGAAACGCTTCCTGCAG**TGA**TGGTGGAGGGGTCCTGAT
GGCTCTGTG

The disclosed GPCR1c protein having 316 amino acid residues is presented using the one-letter code in Table 1J. An analysis using the PSORT program predicts that the ba113a10_C protein localizes in the plasma membrane with a certainty=0.6400; it may also be localized in the Golgi body with a moderate certainty. It is also predicted that protein has a signal peptide whose most likely cleavage site is between residues 44 and 45: GNG-VL, indicated by a slash in Table 1J.

Table 1J. Encoded GPCR1c protein sequence (SEQ ID NO:6).
MVSSNQTSPVLGFLLLGLSAHPKLEKTFFVLILLMYLVILLGNG/VLILVTILDSRLDTPMYFFLGN LSFLDICYTTSSSLTASLTPRKTISFSACAVQMFLSLAMGATECVLLSMMAFDRYVAICNPLWYPEVMNK ATYVPMAAGSWVAGSLTAMVQTPLALRLPFCGDNIINHFTCEILAVLKLACADISVNVISMGVANVIFLG VPVLFISFSYVFIIATILRIPSAEGRKKAFSTCSAHLTVVIVFYGTILFMYGKPKSKDPLGADKQDLADK LISLFYGVVTPMLNPIIYSLRNKEVKAAVRNLVFQKRFLQ

A BLASTX search was performed against public protein databases. The full amino acid sequence of the protein of the invention was found to have 270 of 317 amino acid residues (85%) identical to, and 287 of 317 residues (90%) positive with, a 319 amino acid residue OR from Mus musculus (ptnr:TREMBLNEW-ACC:CAB55596, E= 5.2 e-138). The disclosed GPCR1b protein (SEQ ID NO:6) has good identity with a number of olfactory receptor proteins, as shown in Table 1K. The Patp results include those listed in Table 1L.

Table 1K. BLAST results for GPCR1c

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|11464981|ref|NP_ 062349.1|	OR 37e *Mus musculus*	319	242/317 (76%)	257/317 (80%)	e-120
gi|11276075|ref|NP_ 062346.1|	OR 37a *Mus musculus*	319	241/319 (75%)	256/319 (87%)	e-119
gi|11276077|ref|NP_ 062347.1|	OR 37b *Mus musculus*	318	236/319 (78%)	255/319 (87%)	e-116
gi|11276079|ref|NP_ 062348.1|	OR 37c *Mus musculus*	318	235/308 (76%)	250/308 (80%)	e-116
gi|10092669|ref|NP_ 063950.1|	OR Family 2, Subfamily S, member 2 *Homo sapiens*	309	220/309 (71%)	242/309 (78%)	e-108

Table 1L. Patp alignments of GPCR1c

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob. P(N)
patp:B43266 Human ORFX ORF3030 polypeptide sequence SE...	+2	689	4.4e-67
patp:Y83390 Olfactory receptor protein OLF-5 - H. sapiens...	+2	649	7.5e-63
patp:R27868 Odorant receptor clone F5 - Rattus rattus,...	+2	648	9.6e-63
patp:Y83389 Olfactory receptor protein OLF-4 - H. sapiens...	+2	647	1.2e-62
patp:Y96680 Murine olfactory receptor ligand-binding ...	+2	644	2.6e-62
patp:R27872 Odorant receptor clone I7 - Rattus rattus,..	+2	639	8.7e-62
patp:Y83387 Olfactory receptor protein OLF-2 - H. sapiens...	+2	635	2.3e-61

For example, a BLAST against patp: Y83390, a 305 amino acid olfactory receptor protein from *Homo sapiens*, produced 135/308 (43%) identity, and 201/308 (65%) positives (E = 7.5e-63). PCT publication WO 00/21985.

The amino acids differences between the three GPCR1 proteins are shown in Table 1M. Deletions are marked by a delta (Δ). The differences between the three proteins appear to be localized to a few distinct regions. Thus, these proteins may have similar functions, such as serving as olfactory or chemokine receptors (see below).

Table 1M. Differences for GPCR1 Proteins

Position	4	8	9	11	12	14	49	51	52	53	55	57	58
GPCR1a	A	A	S	T	E	I	M	V	S	N	H	H	M
GPCR1b	A	A	S	T	E	I	M	V	S	N	H	H	M
GPCR1c	S	S	P	L	G	L	V	I	L	D	R	D	T
Position	79	80	81	82	84	85	86	106	132	135	138	141	143
GPCR1a	V	P	L	I	D	S	F	F	R	V	S	A	M
GPCR1b	V	P	L	I	D	S	F	F	R	V	S	A	M
GPCR1c	S	Δ	Δ	Δ	T	A	S	L	W	E	N	T	V
Position	145	146	156	157	158	163	166	204	251	252	304	311	313
GPCR1a	H	K	T	A	S	S	M	T	V	I	D	D	I
GPCR1b	Δ	I	T	A	S	S	M	T	V	I	D	D	I
GPCR1c	Δ	M	L	T	A	P	L	A	I	V	E	N	V
Position	317	319	320										
GPCR1a	C	A	Δ										
GPCR1b	C	A	Δ										
GPCR1c	R	L	Q										

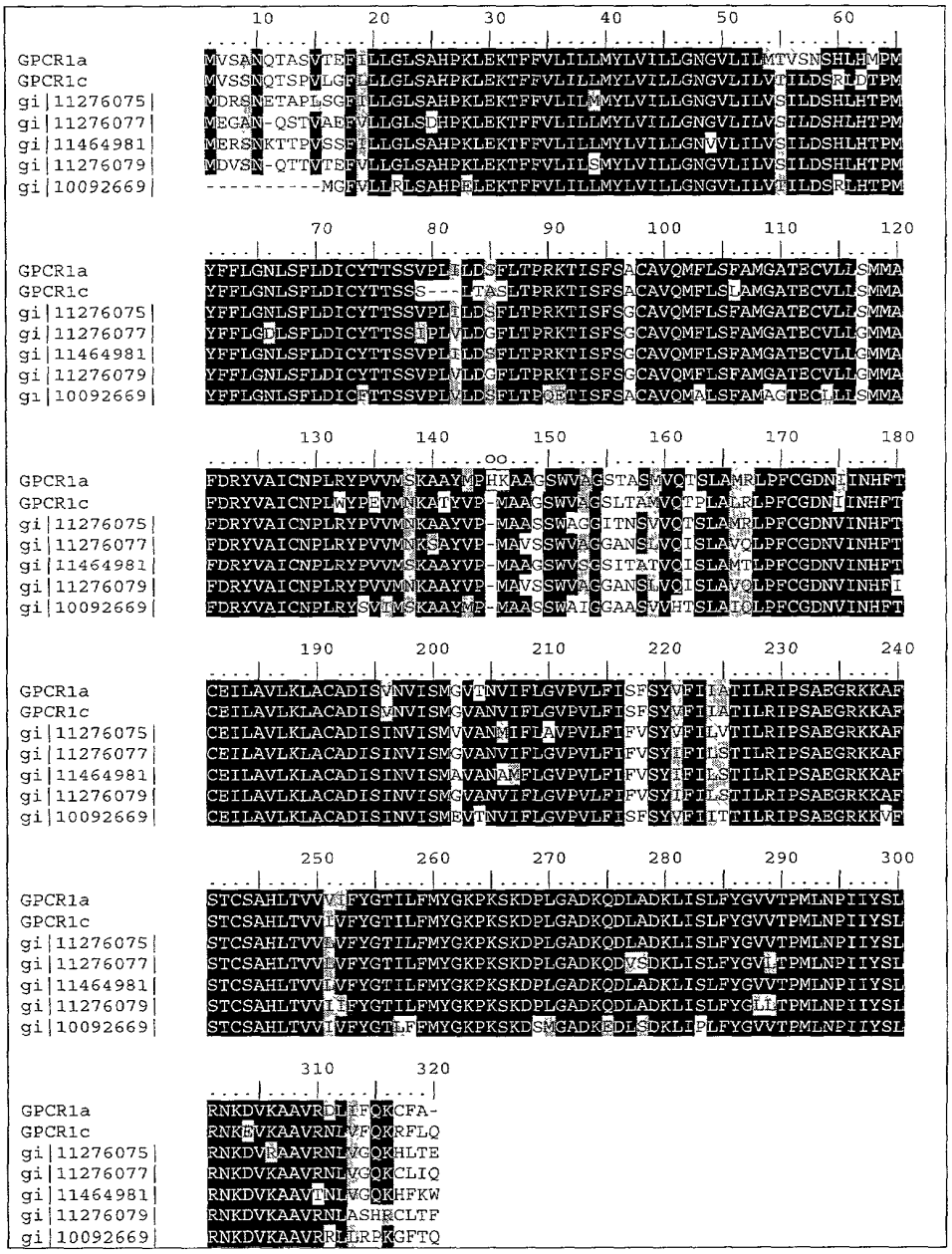
A ClustalW analysis comparing disclosed proteins of the invention with related OR protein sequences is given in Table 1N, with GPCR1a shown on line 1, and GPCR1c on line 2.

In the ClustalW alignment of the GPCR1a protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be mutated to a much broader extent without altering protein structure or function. Unless specifically addressed as GPCR1a GPCR1b, or GPCR1c, any reference to GPCR1 is assumed to encompass all variants. Residue differences between any GPCRX variant sequences herein are written to show the residue in the "a" variant and the residue position with respect to the "a" variant. GPCR residues in all following sequence alignments that differ between the individual GPCR variants are highlighted with a box and marked with the (o) symbol above the variant residue in all alignments herein. For example, the protein shown in line 1 of Table 1N depicts the sequence for GPCR1a, and the positions where GPCR1b differs are marked with a (o) symbol and are highlighted with a box. All GPCR1 proteins have significant homology to olfactory receptor

(OR) proteins: 37a, 37b, 37e, and 37c from *Mus musculus* and to a human OR protein member 2 from family 2, subfamily S (see also Tables 1C, 1G, and 1K).

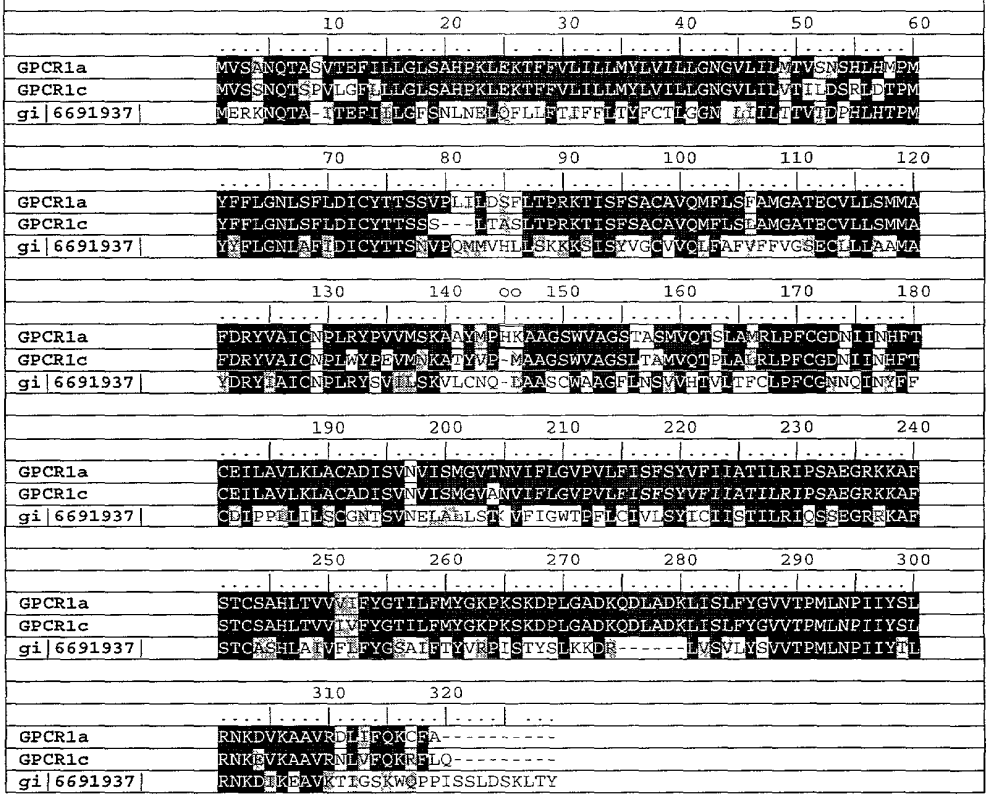
Table 1N. ClustalW Analysis of GPCR1

1) Novel GPCR1a (SEQ ID NO 2)
2) Novel GPCR1c (SEQ ID NO 6)
3) gi|11276075|ref|NP_062346 1| olfactory receptor 37a *Mus musculus* (SEQ ID NO 39)
4) gi|11276077|ref|NP_062347 1| olfactory receptor 37b *Mus musculus* (SEQ ID NO 40)
5) gi|11464981|ref|NP_062349 1| olfactory receptor 37e *Mus musculus* (SEQ ID NO 41)
6) gi|11276079|ref|NP_062348 1| olfactory receptor 37c *Mus musculus* (SEQ ID NO 42)
7) gi|10092669|ref|NP_063950 1| OR, fam 2, subfam S, mem 2 *H sapiens* (SEQ ID NO 43)



The GPCR1 proteins also have regions of identity with a 321 amino acid human transmembrane receptor (gi|6691937|emb|CAB65797.1| bA150A6.2, novel 7 transmembrane receptor (rhodopsin family), olfactory receptor- like protein (hs6M1-21)), as shown in line 3 of in Table 1O (SEQ ID NO: 44).

Table 1O. Alignment of GPCR1a & c with SEQ ID NO:44



The presence of identifiable domains in GPCR1, as well as all other GPCRX proteins, was determined by searches using software algorithms such as PROSITE, DOMAIN, Blocks, Pfam, ProDomain, and Prints, and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (http:www.ebi.ac.uk/ interpro). DOMAIN results, *e.g.*, for GPCR1a as disclosed in Table 1P, were collected from the Conserved Domain Database (CDD) with Reverse Position Specific BLAST analyses. This BLAST analysis software samples domains found in the Smart and Pfam collections. For Table 1P and all successive DOMAIN sequence alignments, fully conserved single residues are indicated by black shading and "strong" semi-conserved residues are indicated by grey shading. The "strong" group of conserved amino acid residues may be any one of the following groups of amino acids: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW.

Table 1P lists the domain description from DOMAIN analysis results against GPCR1a. The region from amino acid residue 53 through 239 (SEQ ID NO:2) most probably ($E = 2e^{-19}$) contains a "seven transmembrane receptor (rhodopsin family) fragment" domain, aligned here with residues 12-180 of the 7tm_1 entry (TM7, SEQ ID NO:45, see Table 1Q for the complete sequence) of the Pfam database. This indicates that the GPCR1 sequence has properties similar to those of other proteins known to contain this domain as well as to the 377 amino acid 7tm domain itself. GPCR1a also has identity to another region of the TM7 protein. The region from amino acid residue 226 through 298 (of SEQ ID NO:2) aligns with amino acid residues 310-377 of TM7 ($E = 3e^{-4}$). GPCR1b and GPCR1c also align to this domain: residues 53-238 of GPCR1b align with residues 12-180 of TM7 ($E = 6e^{-19}$) and residues 225-297 of GPCR1b align with residues 310-377 of TM7 ($E = 3e^{-4}$); residues 55-235 of GPCR1c align with residues 14-180 of TM7 ($E = 2e^{-12}$) and residues 222-294 of GPCR1c align with residues 310-377 of TM7 ($E = 2e^{-4}$).

Table 1P. Domain Analysis of GPCR1a

Sbjct: 7 transmembrane receptor (rhodopsin family) fragment
Gnl|Pfam|pfam00001; Length = 377
Score = 89.7 bits (221), Expect = 2e-19

```
GPCR1a  ----------NSHLHMPMYFFLGNLSFLDICYTTSSVPLLLDSFLTPRKTISFSACAVQ
TM7     GNVLVCMAVSREKALQTTTNYLIVSLAVADLLVATLVMPWVVYLEVVGEWKFSRIHCDIF

GPCR1a  MFLSFAMGATECVLLSMMAFDRYVAICNPLRYPVVMSKAAYMPHKAAGSWVAGSTASMVQ
TM7     VTLDVMMCTASILNLCAISIDRYTAVAMPMLYNTRYSSKRRVTVMIAIVWVLSFTISCPM

GPCR1a  TSLAMRLPFCGDNIINHFTCELLAVLKLACADISVNVISMGVTNVIFLGVPVLFISFSYV
TM7     LFGLNNTDQNE--------CIIANP--------AFVVYSSIVSFYV----PFIVTLLVYI

GPCR1a  FIIATILR--IPSAEGRKKA----------------------------------------
TM7     KIYIVLRRRRKRVNTKRSSRAFRANLKAPLKGNCTHPEDMKLCTVIMKSNGSFPVNRRRV
```

The representative member of the 7 transmembrane receptor family is the D2 dopamine receptor from *Bos taurus* (SWISSPROT: locus D2DR_BOVIN, accession P20288; gene index 118205). The D2 receptor is an integral membrane protein and belongs to Family 1 of G-protein coupled receptors. The activity of the D2 receptor is mediated by G proteins which inhibit adenylyl cyclase. Chio *et al.*, Nature 343:255-269 (1990). Residues 51-427 of this 444 amino acid protein are considered to be the representative TM7 domain, shown in Table 1Q.

Table 1Q Amino Acid sequence for TM7 (SEQ ID NO:45)
GNVLVCMAVSREKALQTTTNYLIVSLAVADLLVATLVMPWVVYLEVVGEWKFSRIHCDIF VTLDVMMCTASILNLCAISIDRYTAVAMPMLYNTRYSSKRRVTVMIAIVWVLSFTISCPM LFGLNNTDQNECIIANPAFVVYSSIVSFYVPFIVTLLVYIKIYIVLRRRRKRVNTKRSSR AFRANLKAPLKGNCTHPEDMKLCTVIMKSNGSFPVNRRRVEAARRAQELEMEMLSSTSPP ERTRYSPIPPSHHQLTLPDPSHHGLHSTPDSPAKPEKNGHAKTVNPKIAKIFEIQSMPNG KTRTSLKTMSRRKLSQQKEKKATQMLAIVLGVFIICWLPFFITHILNIHCDCNIPPVLYS AFTWLGYVNSAVNPIIY

The 7 transmembrane receptor family includes a number of different proteins, including, for example, serotonin receptors, dopamine receptors, histamine receptors, andrenergic receptors, cannabinoid receptors, angiotensin II receptors, chemokine receptors, opioid receptors, G-protein coupled receptor (GPCR) proteins, olfactory receptors (OR), and the like. Some proteins and the Protein Data Base Ids/gene indexes include, for example: rhodopsin (129209); 5-hydroxytryptamine receptors; (112821, 8488960, 112805, 231454, 1168221, 398971, 112806); G protein-coupled receptors (119130, 543823, 1730143, 132206, 137159, 6136153, 416926, 1169881, 136882, 134079); gustatory receptors (544463, 462208); c-x-c chemokine receptors (416718, 128999, 416802, 548703, 1352335); opsins (129193, 129197, 129203); and olfactory receptor-like proteins (129091, 1171893, 400672, 548417);

Expression information for GPCRX RNA was derived using tissue sources including, but not limited to, proprietary database sources, public EST sources, literature sources, and/or RACE sources, as described in the Examples.

The nucleic acids and proteins of GPCR1 are useful in potential therapeutic applications implicated in various GPCR- or olfactory receptor (OR)-related pathologies and/or disorders. For example, a cDNA encoding the G-protein coupled receptor-like protein may be useful in gene therapy, and the G-protein coupled receptor-like protein may be useful when administered to a subject in need thereof. The novel nucleic acid encoding GPCR1 protein, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. The GPCRX nucleic acids and proteins are useful in potential diagnostic and therapeutic applications implicated in various diseases and disorders described below and/or other pathologies. For example, the compositions of the present invention will have efficacy for treatment of patients suffering from: cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, transplantation,

adrenoleukodystrophy, congenital adrenal hyperplasia, fertility, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, graft versus host disease, bronchial asthma, and other diseases, disorders and conditions of the like. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from neoplasm, adenocarcinoma, lymphoma, prostate cancer, uterus cancer, immune response, AIDS, asthma, Crohn's disease, multiple sclerosis, and Albright Hereditary Ostoeodystrophy. Additional GPCR-related diseases and disorders are mentioned throughout the Specification.

Further, the protein similarity information, expression pattern, and map location for GPCR1 suggests that GPCR1 may have important structural and/or physiological functions characteristic of the GPCR family. Therefore, the nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), and (v) a composition promoting tissue regeneration *in vitro* and *in vivo* (vi) biological defense weapon.

These materials are further useful in the generation of antibodies that bind immuno-specifically to the novel GPCR1 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-GPCRX Antibodies" section below. The disclosed GPCR1 protein has multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated GPCR1 epitope is from about amino acids 75 to 90. In another embodiment, a GPCR1 epitope is from about amino acids 125 to 160. In additional embodiments, GPCR1 epitopes are from amino acids 225 to 250, 260-280, and from amino acids 290 to 325. These novel proteins can also be used to develop assay system for functional analysis.

GPCR2

An additional GPCR-like protein of the invention, referred to herein as GPCR2a, is an Olfactory Receptor ("OR")-like protein. The novel nucleic acid of 1254 nucleotides (11612531_1, SEQ ID NO:7) encoding a novel G-protein coupled receptor-like protein is shown in Table 2A. SeqCalling fragments for GPCR2a came from placenta, indicating that it may be expressed in tissues important for female health.

Table 2A. GPCR2a Nucleotide Sequence (SEQ ID NO:7)

```
ACGCGCTTCGACATACTATTCCTGGTGGGCTCTTCCTACATATCAGGTCGTTAAATAAGCTGCCAGATTTCTGCCTTTACAGCCCAA
GGAGCTTGTCATGGACCATGGGCATGGAGGGTCTTCTCCAGAACTCCACTAACTTCGTCCTCACAGGCCTCATCACCCATCCTGCCT
TCCCCGGGCTTCTCTTTGCAATAGTCTTCTCCATCTTTGTGGTGGCTATAACAGCCAACTTGGTCATGATTCTGCTCATCCACATGG
ACTCCCGCCTCCACACACCCATGTACTTCTTGCTCAGCCAGCTCTCCATCATGGATACCATCTACATCTGTATCACTGTCCCCAAGA
TGCTCCAGGACCTCCTGTCCAAGGACAAGACCATTTCCTTCATGGGCTGTGCAGTTCAGATCTTCCTCTACCTGACCCTGATTGGAG
GGGAATTCTTCCTGCTGGGTCTCATGGCCTATGACCGCTATGTGGCTGTGTGCAACCCTCTACGGTACCCTCTCCTCATGAACCGCA
GGGTTTGCTTATTCATGGTGGTCGGCTCCTGGGTTGGTGGTTCCTTGGATGGGTTCATGCTGACTCCTGTCACTATGAGTTTCCCCT
TCTGTAGATCCCGAGAGATCAATCACTTTTTCTGTGAGATCCCAGCCGTGCTGAAGTTGTCTTGCACAGACACGTCACTCTATGAGA
CCCTGATGTATGCCTGCTGCGTGCTGATTATCCCTCTATCTGTCATCTCTGTGTCCTACACGCACATCCTCCTGACTGTCCACAGGA
TGAACTCTGCTGAGGGCCGGCGCAAAGCCTTTGCTACGTGTTCCTCCCACATTATGGTGGTGAGCGTTTTCTACGGGGCAGCTTTCT
ACACCAACGTGCAGCCCCACTCCTACCACACTCCAGAGAAAGATAAAGTGGTGTCTGCCTTCTACACCATCCTCACCCCCATGCTCA
ACCCACTCATCTACAGCTTGAGGAATAAAGATGTGGCTGCAGCTCTGAGGAAAGTACTAGGGAGATGTGGCTCCTCCCAGAGCATCA
GGGTGATGACTGTGTGATCAGGAAGGACTAGCAGGGACTCCCAGAGTATCAGCGTGGTGACTATGATCAGGAAGGACTAGTGGGGAC
TCCTAGAGCATCAGGGTGGCGACTGTGATCAGGAAGGACTAGCAAGGACTAGCGCAAACATCTGCGGTGCTGCGGCCAATAACGCAG
CTATTACAGAAAATATGTTATTGGTTCTGAAGAAGT
```

An open reading frame (ORF) for GPCR2a was identified from nucleotides 105 to 1058. The disclosed GPCR2a polypeptide (SEQ ID NO:8) encoded by SEQ ID NO:7 is 318 amino acid residues and is presented using the one-letter code in Table 2B. The GPCR2a protein was analyzed for signal peptide prediction and cellular localization. SignalPep results predict that GPCR2a is cleaved between position 42 and 43 of SEQ ID NO:8, *i.e.*, at the slash in the amino acid sequence ITA-NL. Psort and Hydropathy profiles also predict that GPCR2a contains a signal peptide and is likely to be localized at the plasma membrane (certainty of 0.6400).

Table 2B. Encoded GPCR2a protein sequence (SEQ ID NO:8).

```
MGMEGLLQNSTNFVLTGLITHPAFPGLLFAIVFSIFVVAITA/NLVMILLIHMDSRLHTPMYFLLSQLS
IMDTIYICITVPKMLQDLLSKDKTISFMGCAVQIFLYLTLIGGEFFLLGLMAYDRYVAVCNPLRYPLLM
NRRVCLFMVVGSWVGGSLDGFMLTPVTMSFPFCRSREINHFFCEIPAVLKLSCTDTSLYETLMYACCVL
IIPLSVISVSYTHILLTVHRMNSAEGRRKAFATCSSHIMVVSVFYGAAFYTNVQPHSYHTPEKDKVVSA
FYTILTPMLNPLIYSLRNKDVAAALRKVLGRCGSSQSIRVMTV
```

The full amino acid sequence of the protein of the invention was found to have 151 of 216 amino acid residues (69%) identical to, and 177 of 216 residues (81%) positive with, a 216 amino acid residue olfactory receptor from *Homo sapiens* (ptnr:SPTREMBL-ACC: O43869) (E = 2.2 e^{-61}). The protein encoded by GPCR2 (SEQ ID NO:7) has significant homology to olfactory, odorant, and taste chemoreceptors and belongs to the family of G-Protein coupled receptors (GPCRs). This family of genes has been used as a target for small molecule drugs and GPCRs are expressed on the plasma membrane and are also a suitable target for protein drugs like therapeutic antibodies, cytotoxic antibodies and diagnostic antibodies.

As shown in Table 2C, BLAST analysis shows that GPCR2a has significant homology with a number of olfactory receptors. The Patp results include those listed in Table 2D.

Table 2C. GPCR2a BLAST results

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob. P(N)
Ptnr:SPTREMBL-ACC:O43869 OLFACTORY RECEPTOR - HOMO SAPIENS...	827	1.8e-82	1
Ptnr:SWISSPROT-ACC:P23275 OLFACTORY RECEPTOR 15 (OR3) - M...	712	2 7e-70	1
Ptnr:SPTREMBL-ACC:Q90808 OLFACTORY RECEPTOR 4 - GALLUS GALLUS...	708	7.3e-70	1
Ptnr:TREMBLNEW-ACC.CAB55593 OLFACTORY RECEPTOR - Mus musculus...	706	1.2e-69	1
Ptnr:TREMBLNEW-ACC:CAB55594 OLFACTORY RECEPTOR - Mus musculus...	704	1.9e-69	1
Ptnr:SPTREMBL-ACC:Q63394 OL1 RECEPTOR - RATTUS NORVEGICUS...	702	3.1e-69	1
Ptnr:TREMBLNEW-ACC:CAB55592 OLFACTORY RECEPTOR - Mus musculus...	698	8.3e-69	1
Ptnr:TREMBLNEW-ACC:CAB55596 OLFACTORY RECEPTOR - Mus musculus...	697	1.1e-68	1
Ptnr:SWISSPROT-ACC.Q95156 OLFACTORY RECEPTOR-LIKE PROTEIN...	695	1 7e-68	1
Ptnr:SWISSPROT-ACC:Q13606 OLFACTORY RECEPTOR-LIKE PROTEIN...	694	2.2e-68	1
Ptnr:SWISSPROT-ACC.Q13607 OLFACTORY RECEPTOR-LIKE PROTEIN...	689	7.5e-68	1
Ptnr:TREMBLNEW-ACC:AAC64376 OLFACTORY RECEPTOR-LIKE PROTEIN...	685	2.0e-67	1

Table 2D. Patp alignments of GPCR2a

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob. P(N)
patp:W75960 Human olfactory OLRCC15 receptor - H. sapiens.	+3	862	2.0e-85
patp:Y96676 Murine olfactory receptor ligand-binding r...	+3	860	3.3e-85
patp:Y83390 Olfactory receptor protein OLF-5 - H. sapiens...	+3	691	2.7e-67
patp:Y96668 Murine olfactory receptor ligand-binding r...	+3	677	8.1e-66
patp:Y96666 Murine olfactory receptor ligand-binding r...	+3	672	2.8e-65
patp:Y96670 Murine olfactory receptor ligand-binding r..	+3	672	2.8e-65
patp:Y83389 Olfactory receptor protein OLF-4 - H. sapiens...	+3	670	4.5e-65

For example, a BLAST against patp: W75960, a 316 amino acid olfactory receptor protein from *Homo sapiens*, produced 164/311 (52%) identity, and 217/311 (69%) positives (E = 2.0e-85). See, European patent EP 867508, disclosing a novel olfactory receptor of the 7-transmembrane receptor family. Also, a BLAST against Y96676, a 223 amino acid olfactory receptor from *Mus musculus*, produced 156/223 (69%) identity, and 184/223 (82%) positives (E = 3.3e-85, SEQ ID NO:87). WO 00/35274. These proteins each show long segments of amino acid identity, as shown in Tables 2E and 2F, respectively.

Table 2E: Alignment of GPCR2a withW75960 Human OR OLRCC15 (SEQ ID NO:86)

Length = 316 Plus Strand HSPs:
Score = 862 (303.4 bits), Expect = 2.0e-85, P = 2.0e-85
Identities = 164/311 (52%), Positives = 217/311 (69%), Frame = +3

```
                10        20        30        40        50        60
                ....|....|....|....|....|....|....|....|....|....|....|
GPCR2a  MGMENGLLQNSTNEVLTGLITEPAFPGLLFAIVFSIFVVAITANLVMILLIHMDSRLHTPM 60
W75960  ---ENSTFNS-DFIFLGIFNESPTHTFLFFLVLAIFSVAFMGNSVMVLLIYLDTQLHTPM 56

                70        80        90        100       110       120
                ....|....|....|....|....|....|....|....|....|....|....|
GPCR2a  YFLLSQLSIMDTIYICITVPKMLQDLLSKDKTISFMGCAVQIFLYLSLIGGEFFLLGLMA 120
W75960  YLLLSQLSLMDLMLICTTVPKMAFNYLSGSKSISMAGCATQIFFYTSLLGSECFLLAVMA 116
```

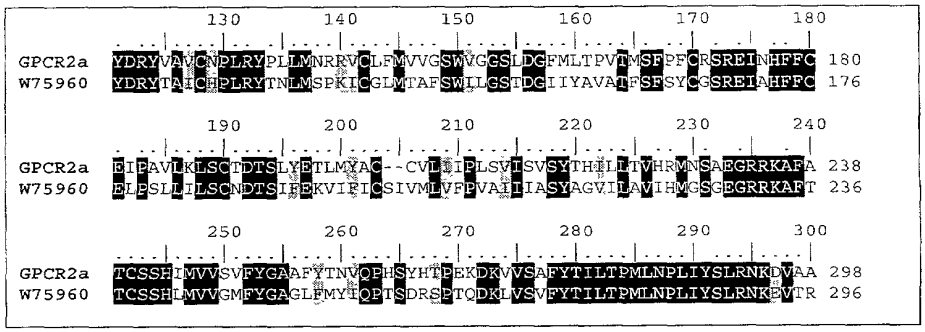
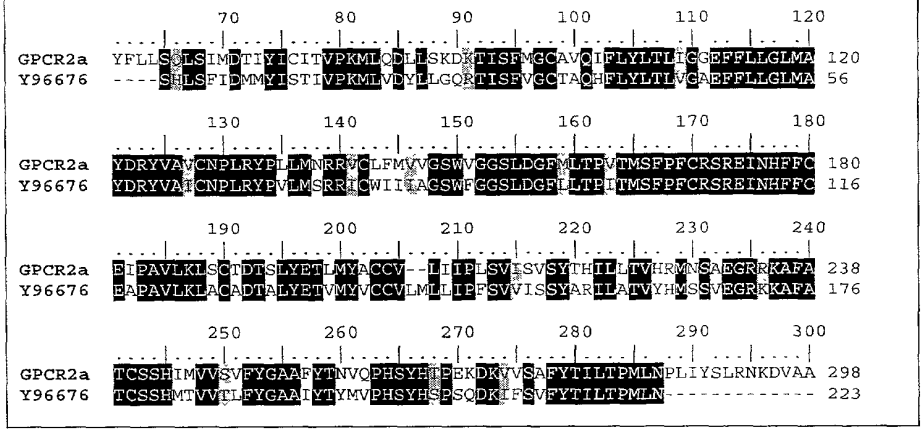


Table 2F: Alignment of GPCR2a Y96676 Murine OR ligand-binding region 15 (SEQ ID NO:87)

Length = 223 Plus Strand HSPs:
Score = 860 (302.7 bits), Expect = 3.3e-85, P = 3.3e-85
Identities = 156/223 (69%), Positives = 184/223 (82%), Frame = +3



Other BLAST results including the sequences used for ClustalW analysis is presented in Table 2G.

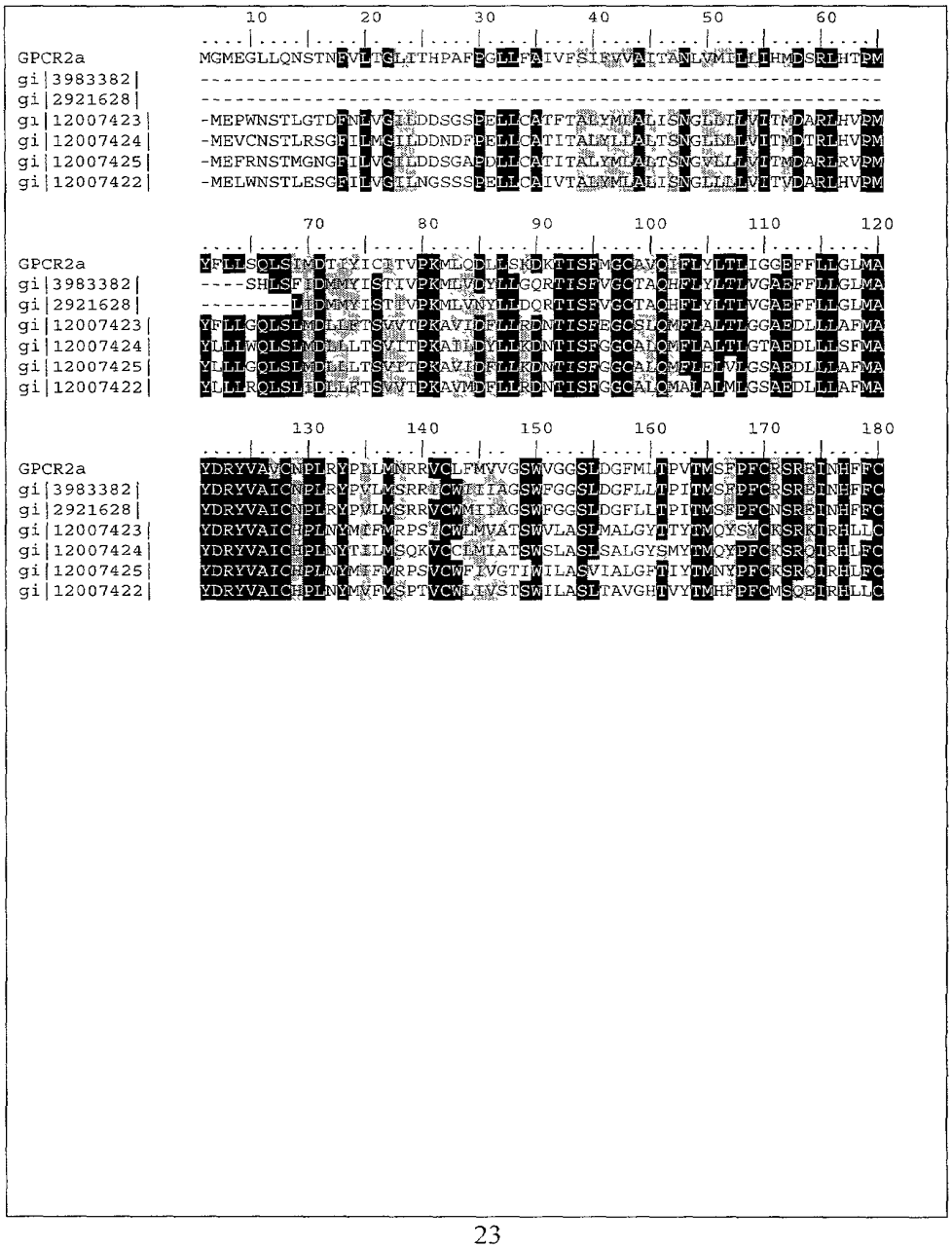
Table 2G. BLAST results for GPCR2a

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
Gi|3983382|gb|AAD13319.1| (AF102527)	OR E3 *Mus musculus*	223	156/223 (69%)	184/223 (81%)	2e-78
Gi|2921628|gb|AAC39611.1| (U86215)	OR *Homo sapiens*	216	151/216 (69%)	177/216 (81%)	5e-75
Gi|12007423|gb|AAG45196.1| (AF321234)	T2 OR *Mus musculus*	316	155/298 (52%)	205/298 (68%)	2e-72
Gi|12007424|gb|AAG45197.1| (AF321234)	T3 OR *Mus musculus*	315	156/304 (51%)	209/304 (68%)	2e-72
Gi|12007425|gb|AAG45198.1| (AF321234)	T4 OR *Mus musculus*	319	152/304 (50%)	207/304 (68%)	9e-72
Gi|12007422|gb|AAG45195.1| (AF321234)	T1 OR *Mus musculus*	316	156/305 (51%)	207/305 (67%)	2e-67

This information is presented graphically in the multiple sequence alignment given in Table 2H (with GPCR2a being shown on line 1) as a ClustalW analysis comparing GPCR2a with related protein sequences.

Table 2H. Information for the ClustalW proteins:

1) Novel GPCR2a (SEQ ID NO 8)
2) gi|3983382|gb|AAD13319 1| olfactory receptor E3 *Mus musculus* (SEQ ID NO 46)
3) gi|2921628|gb|AAC39611 1| olfactory receptor *Homo sapiens* (SEQ ID NO 47)
4) gi|12007423|gb|AAG45196 1| T2 olfactory receptor *Mus musculus* (SEQ ID NO 48)
5) gi|12007424|gb|AAG45197 1| T3 olfactory receptor *Mus musculus* (SEQ ID NO 49)
6) gi|12007425|gb|AAG45198 1| T4 olfactory receptor *Mus musculus* (SEQ ID NO 50)
7) gi|12007422|gb|AAG45195 1| T1 olfactory receptor *Mus musculus* (SEQ ID NO 51)



GPCR2b

The target sequence identified as GPCR2a (Accession Number 11612531_1) was subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on in silico predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences sequences from other species. These primers were then employed in PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provided GPCR2b, the sequence reported below in tables 2I and 2J, and designated Accession Number 11612531_1_da1. The nucleotide sequence for the 1014 nucleotide sequence for GPCR2b is shown in Table 2I and is referred to by SEQ ID NO:84.

Table 2I. GPCR2b Nucleotide Sequence (SEQ ID NO:84)

CAGCCCAAGGAGCTTGTCATGGACCATGGGCATGGAGGGTCTTCTCCAGAACTCCACTAACTTCGTCCTC
ACAGGCCTCATCACCCATCCTGCCTTCCCCGGGCTTCTCTTTGCAATAGTCTTCTCCATCTTTGTGGTGG
CTATAACAGCCAACTTGGTCATGATTCTGCTCATCCACATGGACTCCCGCCTCCACACACCCATGTACTT
CTTGCTCAGCCAGCTCTCCATCATGGATACCATCTACATCTGTATCACTGTCCCCAAGATGCTCCAGGAC
CTCCTGTCCAAGGACAAGACCATTTCCTTCCTGGGCTGTGCAGTTCAAATCTTCCTCTACCTGACCCTGA
TTGGAGGGGAATTCTTCCTGCTGGGTCTCATGGCCTATGACCGCTATGTGGCTGTGTGCAACCCTCTACG
GTACCCTCTCCTCATGAACCGCAGGGTTTGCTTATTCATGGTGGTCGGCTCCTGGGTTGGTGGTTCCTTG
GATGGGTTCATGCTGACTCCTGTCACTATGAGTTTCCCCTTCTGTAGATCCCGAGAGATCAATCACTTTT
TCTGTGAGATCCCAGCCGTGCTGAAGTTGTCTTGCACAGACACGTCACTCTATGAGACCCTGATGTATGC
CTGCTGCGTGCTGATGCTGCTTATCCCTCTATCTGTCATCTCTGTGTCCTACACGCACATCCTCCTGACT
GTCCACAGGATGAACTCTGCTGAGGGCCGGCGCAAAGCCTTTGCTACGTGTTCCTCCCACATTATGGTGG
TGAGCGTTTTCTACGGGGCAGCCTTCTACACCAACGTGCAGCCCCACTCCTACCATACTCCAGAGAAAGA
TAAAGTGGTGTCTGCCTTCTACACCATCCTCACCCCCATGCTCAACCCACTCATCTACAGCTTGAGGAAT
AAAGATGTGGCTGCAGCTCTGAGGAAAGTACTAGGGAGATGTGGCTCCTCCCAGAGCATCAGGGTGATGA
CTGTGTGATCAGGAAGGACTAGCAGGGACTCCCA

GPCR2b differs from GPCR2a (SEQ ID NO:7) by having 79 fewer nucleotides at the 5' end, 166 fewer nucleotides at the 3' end, a 6 nucleotide insertion (GCTGCT) at bp725-730 (numbered according to GPCR2a): and four base substitutions: A390>C; G407>A T872>c; and C905>T.

The disclosed GPCR2b polypeptide (SEQ ID NO:85) encoded by SEQ ID NO:84 is 320 amino acid residues and is presented using the one-letter code in Table 2J. The GPCR2b protein was analyzed for signal peptide prediction and cellular localization. SignalPep results predict that GPCR2b is cleaved between position 42 and 43 of SEQ ID NO:85, *i.e.*, at the slash in the amino acid sequence ITA-NL. Psort and Hydropathy profiles also predict that GPCR2b is likely to be localized at the plasma membrane (certainty of 0.6400).

Table 2J. Encoded GPCR2b protein sequence (SEQ ID NO:85).

MGMEGLLQNSTNFVLTGLITHPAFPGLLFAIVFSIFVVAITA/NLVMILLIHMDSRLHTPMYFLLSQLSIM
DTIYICITVPKMLQDLLSKDKTISF**LG**CAVQIFLYLTLIGGEFFLLGLMAYDRYVAVCNPLRYPLLMNRR
VCLFMVVGSWVGGSLDGFMLTPVTMSFPFCRSREINHFFCEIPAVLKLSCTDTSLYETLMYACCV**LM**LL**I**
PLSVISVSYTHILLTVHRMNSAEGRRKAFATCSSHIMVVSVFYGAAFYTNVQPHSYHTPEKDKVVSAFYT
ILTPMLNPLIYSLRNKDVAAALRKVLGRCGSSQSIRVMTV

GPCR2b differs from GPCR2a at position 96 (M>L) and at position 209 (I>L); additionally, GPCR has a two amino acid insertion, as shown in bold in Table 2J.

The BLASTP (non-redundant composite database), best human BLAST results, and BLASTX (non-redundant) results for GPCR2b are shown in Tables 2K, 2L, and 2M, respectively.

Table 2K. GPCR2b BLASTP

BLASTP: ptnr:SWISSPROT-ACC:P23275 OLFACTORY RECEPTOR 15 (OR3) - Mus musculus (Mouse),312 aa. (SEQ ID NO:90)
Top Previous Match Next Match Length = 312
Score = 719 (253.1 bits), Expect = 6.5e-71, P = 6.5e-71
Identities = 134/300 (44%), Positives = 200/300 (66%)

```
Query:     9 NSTNFVLTGLITHPAFPGLLFAIVFSIFVVAITANLVMILLIHMDSRLHTPMYFLLSQLS 68
             +|  |+| |+  ||    + ||++ + +++ +  || +|||  +|+|||||||| || ||
Sbjct:     8 SSGTFILMGVSDHPHLEIIFFAVILASYLLTLVGNLTIILLSRLDARLHTPMYFFLSNLS 67

Query:    69 IMDTIYICITVPKMLQDLLSKDKTISFLGCAVQIFLYLTLIGGEFFLLGLMAYDRYVAVC 128
              +|  +   +||+||++|   |||||+ ||  |++++| |   |  || +||+|||||||
Sbjct:    68 SLDLAFTTSSVPQMLKNLWGPDKTISYGGCVTQLYVFLWLGATECILLVVMAFDRYVAVC 127

Query:   129 NPLRYPLLMNRRVCLFMVVGSWVGGSLDGFMLTPVTMSFPFCRSREINHFFCEIPAVLKL 188
              || |  +|| |+|  +   ||+||  +  + +  |+  |||  |++++| ||+||++||
Sbjct:   128 RPLHYMTVMNPRLCWGLAAISWLGGLGNSVIQSTFTLQLPFCGHRKVDNFLCEVPAMIKL 187

Query:   189 SCTDTSLYETLMYACCVLMLLIPLSVISVSYTHILLTVHRMNSAEGRRKAFATCSSHIMV 248
             +| |||| | ++    |    ++|+|||  |||   |   | ++ | ||||||| || ||++|
Sbjct:   188 ACGDTSLNEAVLNGVCTFFTVVPVSVILVSYCFIAQAVMKIRSVEGRRKAFNTCVSHLVV 247

Query:   249 VSVFYGAAFYTNVQPHSYHTPEKDKVVSAFYTILTPMLNPLIYSLRNKDVAAALRKVLGR 308
             | +|||+| |  + |       + | +|  ||+++|||+|||||+||||+|  || ++||+
Sbjct:   248 VFLFYGSAIYGYLLPAKSSNQSQGKFISLFYSVVTPMVNPLIYTLRNKEVKGALGRLLGK 307
```

Table 2L. GPCR2b best Human alignment

```
Best Human blast:  ptnr:SPTREMBL-ACC:O95918 DJ271M21.2 (HS6M1-12 (7
TRANSMEMBRANE RECEPTOR(RHODOPSIN FAMILY) (OLFACTORY RECEPTOR LIKE) PROTEIN))
- Homo sapiens (Human), 312 aa.(SEQ ID NO:91)
Top  Previous Match  Next Match            Length = 312
 Score = 713 (251.0 bits), Expect = 2.8e-70, P = 2.8e-70
 Identities = 141/309 (45%), Positives = 203/309 (65%)
Query:      8 QNST-NFVLTGLITHPAFPGLLFAIVFSIFVVAITANLVMILLIHMDSRLHTPMYFLLSQ 66
              |+||  |+| |   ||      || +||+ +++ +  | ++|||  +| +||+|||| ||
Sbjct:      4 QSSTPGFLLLGFSEHPGLERTLFVVVFTSYLLTLVGNTLIILLSALDPKLHSPMYFFLSN 63

Query:     67 LSIMDTIYICITVPKMLQDLLSKDKTISFLGCAVQIFLYLTLIGGEFFLLGLMAYDRYVA 126
              || +|  +    ||+|| +|     |||||| |+|||++|+|   |  || +||+|||||
Sbjct:     64 LSFLDLCFTTSCVPQMLVNLWGPKKTISFLDCSVQIFIFLSLGTTECILLTVMAFDRYVA 123

Query:    127 VCNPLRYPLLMNRRVCLFMVVGSWVGGSLDGFMLTPVTMSFPFCRSREINHFFCEIPAVL 186
              || || |  +++ |+|  +   +|| | ++  + || |+  |||  |+++ | ||+||++
Sbjct:    124 VCQPLHYATIIHPRLCWQLASVAWVIGLVESVVQTPSTLHLPFCPDRQVDDFVCEVPALI 183

Query:    187 KLSCTDTSLYETLMYACCVLMLLIPLSVISVSYTHILLTVHRMNSAEGRRKAFATCSSHI 246
              +||| |||  |  +    | +|++|||+| |||  |   | |+|||+|||||| |||||+
Sbjct:    184 RLSCEDTSYNEIQVAVASVFILVVPLSLILVSYGAITWAVLRINSAKGRRKAFGTCSSHL 243

Query:    247 MVVSVFYGAAFYTNVQPHSYHTPEKDKVVSAFYTILTPMLNPLIYSLRNKDVAAALRKVL 306
               ||++|| +     +|| + +  |+ |     || + || ||||||+||||+|  | |++|
Sbjct:    244 TVVTLFYSSVIAVYLQPKNPYAQERGKFFGLFYAVGTPSLNPLIYTLRNKEVTRAFRRLL 303

Query:    307 GR-CGSSQS 314
              |+  | +||
Sbjct:    304 GKEMGLTQS 312
```

Table 2M. GPCR2b BLASTX

```
BLASTX (Non-redundant):  ptnr:SWISSPROT-ACC:P23275 OLFACTORY RECEPTOR 15
(OR3) - Mus musculus (Mouse),312 aa.(SEQ ID NO:92)
Top  Previous Match  Next Match            Length = 312
  Plus Strand HSPs:
 Score = 719 (253.1 bits), Expect = 3.9e-70, P = 3.9e-70
 Identities = 134/300 (44%), Positives = 200/300 (66%), Frame = +2
Query:     50 NSTNFVLTGLITHPAFPGLLFAIVFSIFVVAITANLVMILLIHMDSRLHTPMYFLLSQLS 229
              +|  |+| |+  ||    + ||++ + +++ +  || +|||  +|+|||||||| || ||
Sbjct:      8 SSGTFILMGVSDHPHLEIIFFAVILASYLLTLVGNLTIILLSRLDARLHTPMYFFLSNLS 67

Query:    230 IMDTIYICITVPKMLQDLLSKDKTISFLGCAVQIFLYLTLIGGEFFLLGLMAYDRYVAVC 409
               +|  +   +||+||++|    |||||+ ||  |++++| |    |  || +||+|||||||
Sbjct:     68 SLDLAFTTSSVPQMLKNLWGPDKTISYGGCVTQLYVFLWLGATECILLVVMAFDRYVAVC 127

Query:    410 NPLRYPLLMNRRVCLFMVVGSWVGGSLDGFMLTPVTMSFPFCRSREINHFFCEIPAVLKL 589
               || |  +|| |+|  +   ||+||  +  + +  |+  |||  |++++| ||+||++||
Sbjct:    128 RPLHYMTVMNPRLCWGLAAISWLGGLGNSVIQSTFTLQLPFCGHRKVDNFLCEVPAMIKL 187

Query:    590 SCTDTSLYETLMYACCVLMLLIPLSVISVSYTHILLTVHRMNSAEGRRKAFATCSSHIMV 769
              +| |||| | ++    |    ++|+|||  |||  |   | ++ | ||||||| || ||++|
Sbjct:    188 ACGDTSLNEAVLNGVCTFFTVVPVSVILVSYCFIAQAVMKIRSVEGRRKAFNTCVSHLVV 247

Query:    770 VSVFYGAAFYTNVQPHSYHTPEKDKVVSAFYTILTPMLNPLIYSLRNKDVAAALRKVLGR 949
              | +|||+| |  + |        + | +| ||+++|||+|||||+||||+|  || ++||+
Sbjct:    248 VFLFYGSAIYGYLLPAKSSNQSQGKFISLFYSVVTPMVNPLIYTLRNKEVKGALGRLLGK 307
```

DOMAIN results for GPCR2 were collected from the Conserved Domain Database (CDD) with Reverse Position Specific BLAST. This BLAST samples domains found in the Smart and Pfam collections. The 7tm_1, a seven transmembrane receptor (rhodopsin family), was shown to have two segments with homology to GPCR2. The region from amino acid

residue 43 through 238 aligns with amino acids 2 through 181 of the "seven transmembrane receptor (rhodopsin family) fragment" domain(SEQ ID NO:45, E= 2e-22), and GPCR2 amino acids 226-289 aligned with residues 313-377 of the 7tm_1 entry (SEQ ID NO:45, E= .008) of the Pfam database. This indicates that the GPCR2 sequence has properties similar to those of other proteins known to contain this domain as well as to the 7tm_1 domain itself.

The disclosed GPCR2 is expressed in tissues that are important in female reproductive health and hence GPCR2 may serve as a drug target for, *e.g.*, premature labor, endometriosis, and *in vitro* fertilization. The homology to the olfactory receptors suggests that an endogenous small molecule ligand regulates this gene and hence drugs structurally similar to the endogenous ligand could serve as agonists and antagonists to regulate the biological effects of GPCR2.

The nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in various GPCR-related pathological disorders and/or OR-related pathological disorders, described further below. For example, a cDNA encoding the olfactory receptor -like protein may be useful in gene therapy, and the olfactory receptor -like protein may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from neoplasm, adenocarcinoma, lymphoma, uterus cancer, immune response, AIDS, asthma, Crohn's disease, multiple sclerosis, and Albright Hereditary Ostoeodystrophy. Other GPCR-related diseases and disorders are contemplated.

The novel nucleic acid encoding the GPCR-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-GPCRX Antibodies" section below. The disclosed GPCR2 protein has multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated GPCR2 epitope is from about amino acids 125 to 135. In another embodiment, a GPCR2 epitope is from about amino acids 225 to 280. In additional embodiments, GPCR2 epitopes are from amino acids 250 to 275, and from amino acids 280 to 310. These novel proteins can also be used to develop assay system for functional analysis.

These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

GPCR3

An additional GPCR-like protein of the invention, referred to herein as GPCR3, is an Olfactory Receptor ("OR")-like protein. The novel nucleic acid was identified on chromosome 6 by TblastN using CuraGen Corporation's sequence file for GPCR probe or homolog, run against the Genomic Daily Files made available by GenBank. The nucleic acid was further predicted by the program GenScan™, including selection of exons. These were further modified by means of similarities using BLAST searches. The sequences were then manually corrected for apparent inconsistencies, thereby obtaining the sequences encoding the full-length protein. The novel nucleic acid of 957 nucleotides (ba145122_B, SEQ ID NO:9) encoding a novel olfactory receptor-like protein is shown in Table 3A. An open reading frame (ORF) was identified beginning with an ATG initiation codon at nucleotides 10-12 and ending with a TGA codon at nucleotides 955-957. Putative untranslated regions upstream from the initiation codon and downstream from the termination codon are underlined in Table 3A, and the start and stop codons are in bold letters.

Table 3A. GPCR3 Nucleotide Sequence (SEQ ID NO:9)

AGATTAGTC**ATG**AAGGCCAACTACAGCGCAGAGGAGCGCTTTCTCCTGCTGGGTTTCTCCGACTGGCCTT
CCCTGCAGCCGGTCCTCTTCGCCCTTGTCCTCCTGTGCTACCTCCTGACCTTGACGGGCAACTCGGCGCT
GGTGCTGCTGGCGGTGCGCGACCCGCGCCTGCACACGCCCATGTACTACTTCCTCTGCCACCTGGCCTTG
GTAGACGCGGGCTTCACTACTAGCGTGGTGCCGCCGCTGCTGGCCAACCTGCGCGGACCAGCGCTCTGGC
TGCCGCGCAGCCACTGCACGGCGCAGCTGTGCGCATCGCTGGCTCTGGGTTCGGCCGAATGCGTCCTCCT
GGCGGTGATGGCTCTGGACCGCGCGGCCGCAGTGTGCCGCCCGCTGCGCTATGCGGGGCTCGTCTCCCCG
CGCCTATGTCGCACGCTGGCCAGCGCCTCCTGGCTAAGCGGCCTCACCAACTCGGTTGCGCAAACCGCGC
TCCTGGCTGAGCGGCCGCTGTGCGCGCCCCGCCTGCTGGGCCACTTCATCTGTGAGCTGCCGGCGTTGCT
CAAGCTGGCCCGCGGAGGCGACGGAGACACTACCGAGAACCAGATGTTCGCCGCCCGCGTGGTCATCCTG
CTGCTGCCGTTTGCCGTCATCCTGGCCTCCTACGGTGCCGTGGCCCGAGCTGTCTGTTGTATGCGGTTCA
GCGGAGGCCGGAGGAGGGCGGTGGGCACGTGTGGGTCCCACCTGACAGCCGTCTGCCTGTTCTACGGCTC
GGCCATCTACACCTACCTGCAGCCCGCGCAGCGCTACAACCAGGCACGGGGCAAGTTCGTATCGCTCTTC
TACACCGTGGTCACACCTGCTCTCAACCCGCTCATCTACACCCTCAGGAATAAGAAAGTGAAGGGGGCAG
CGAGGAGGCTGCTGCGGAGTCTGGGGAGAGGCCAGGCTGGGCAG**TGA**

A putative splice site is located between nucleotides 15 and 16 in SEQ ID NO:9. In one embodiment, nucleotides 1-15 come from exon 1 and nucleotides 16-957 are from exon 2.

The disclosed ba145122_B nucleic acid sequence has 592 of 934 bases (63%) identical to a GPCR mRNA (GENBANK-ID: HUMORLMHC| acc: L35475) from *Homo sapiens* (E= 2.4 e^{-51}).

The disclosed GPCR3 polypeptide (SEQ ID NO:10) encoded by SEQ ID NO:9 is 315 amino acid residues and is presented using the one-letter code in Table 3B. The first 70 amino acids of the disclosed GPCR3 protein were analyzed for signal peptide prediction and cellular localization. SignalP results predict that GPCR3 is cleaved between position 46 and 47 of SEQ

ID NO:10, *i.e.*, at the slash in the amino acid sequence NSA-LV. Psort and Hydropathy profiles also predict that GPCR3 contains a signal peptide and is likely to be localized at the plasma membrane (certainty of 0.6000).

Table 3B. Encoded GPCR3 protein sequence (SEQ ID NO:10).

```
MKANYSAEERFLLLGFSDWPSLQPVLFALVLLCYLLTLTGNSA/LVLLAVRDPRLHTPMYYFLCHLALVDA
GFTTSVVPPLLANLRGPALWLPRSHCTAQLCASLALGSAECVLLAVMALDRAAAVCRPLRYAGLVSPRLC
RTLASASWLSGLTNSVAQTALLAERPLCAPRLLGHFICELPALLKLARGGDGDTTENQMFAARVVILLLP
FAVILASYGAVARAVCCMRFSGGRRRAVGTCGSHLTAVCLFYGSAIYTYLQPAQRYNQARGKFVSLFYTV
VTPALNPLIYTLRNKKVKGAARRLLRSLGRGQAGQ
```

A BLASTX search was performed against public protein databases. The full amino acid sequence of the protein of the invention was found to have 253 of 308 amino acid residues (82%) identical to, and 264 of 308 residues (85%) positive with, the 309 amino acid residue MM17M1-6, 7 transmembrane receptor (OR-like protein) from *Mus musculus* (ptnr:SPTREMBL-ACC:Q9WV09, SEQ ID NO:52) ($E = 1.5\ e^{-128}$). The alignment of these proteins is shown in Table 3C.

Table 3C. Alignment of GPCR3 with Q9WV09 (SEQ ID NO:52).

```
                10        20        30        40        50        60
GPCR3       MKANYSAEERFLLLGFSDWPSLQPVLFALVLLCYLLTLTGNSALVLLAVRDPRLHTPMYY 60
Q9WV09      --ANHSSAERFLLLGFSDWPSLQPVLFALVLLCYLLTLTGNAALVLLAIRDPRLHTPMYY 58

                70        80        90       100       110       120
GPCR3       FLCHLALVDAGFTTSVVPPLLANLRGPALWLPRSHCTAQLCASLALGSAECVLLAVMALD 120
Q9WV09      FLCHLALVDVGFTTSVVPPLLASLRGSMLQLPRAGCMAQLCSSLALGSAECVLLAVMALD 118

               130       140       150       160       170       180
GPCR3       RAAAVCRPLRYAGLVSPRLCRTLASASWLSGLTNSVAQTALLAERPLCAPRLLGHFICEL 180
Q9WV09      RAAAVCNPLRYTSLASPLLCRTLAGVSWLGGLANSAAQTALLAARPLCAPRCLDHFICEL 178

               190       200       210       220       230       240
GPCR3       PALLKLARGSDGDTTENQMFAARVVILLLPFAVILASYGAVARAVCCMRFSGGRRRAVGT 240
Q9WV09      PALLQLACRGGRSATERQMFAARVVILLVPSAVILASYIAVGRAVWGMHSSSGWRKAAST 238

               250       260       270       280       290       300
GPCR3       CGSHLTAVCLFYGSAIYTYLQPAQRYNQARGKFVSLFYTVVTPALNPLIYTLRNKKVKGA 300
Q9WV09      CGSHLTAVCLFYGSAIYTYLQPTHSYNQGRGKFVSLFYTVVTPALNPLIYTLRNKEVKGA 298

               310
GPCR3       ARRLLRSLGRGQAGQ 315
Q9WV09      ALRLLRSLGRP---- 309
```

Patp results include those listed in Table 3D.

Table 3D. Patp alignments of GPCR3

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob. P(N)
patp:B43266 Human ORFX ORF3030 polypeptide sequence SE.	.. +1	703	1.4e-68
patp:B42796 Human ORFX ORF2560 polypeptide sequence SE..	. +1	699	3.8e-68
patp:W21662 Rat spermatid chemoreceptor D-2 - Rattus s..	. +1	619	1.1e-59
patp:W21663 Rat spermatid chemoreceptor D-7 - Rattus s..	. +1	614	3.9e-59
patp:W21664 Rat spermatid chemoreceptor D-8 - Rattus s..	. +1	604	4.4e-58
patp:W21665 Rat spermatid chemoreceptor D-9 - Rattus s...	+1	597	2.4e-57
patp:W75960 Human olfactory OLRCC15 receptor - H. sapiens.	.. +1	579	2.0e-55

The disclosed GPCR3 protein (SEQ ID NO:10) also has good identity with a number of olfactory receptor proteins, as shown in Table 3E.

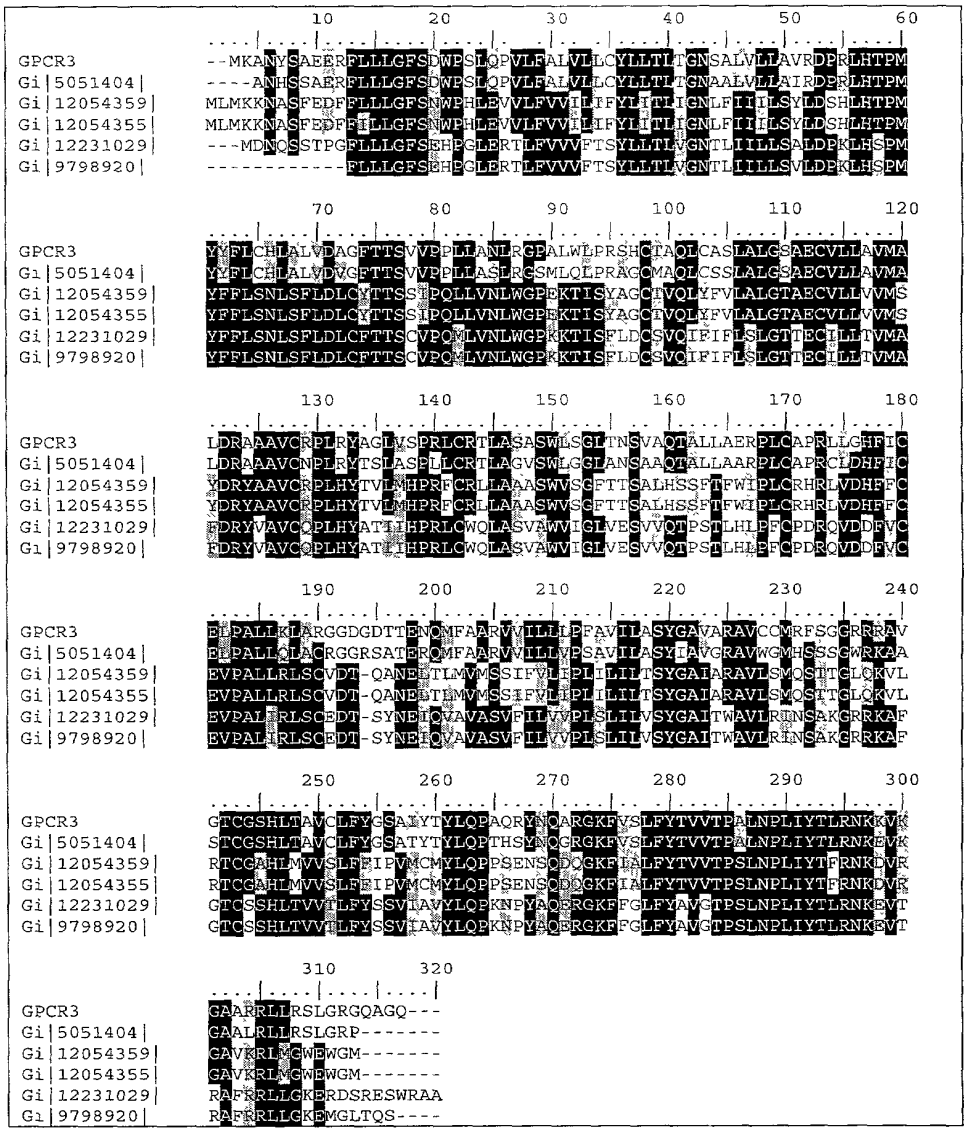
This information is presented graphically in the multiple sequence alignment given in Table 3F (with GPCR3 being shown on line 1) as a ClustalW analysis comparing GPCR3 with related protein sequences.

Table 3E. BLAST results for GPCR3

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
Gi|5051404|emb|CAB45012.1| (AL078630)	573K1.15 - mm17M1-6 (novel 7 transmembrane receptor) *Mus musculus*	309	229/296 (77%)	240/296 (80%)	e-108
Gi|12054359|emb|CAC20487.1| (AJ302567) (AJ02568)	OR *Homo sapiens*	312	145/298 (48%)	191/298 (63%)	2e-69
Gi|12054355|emb|CAC20485.1| (AJ302565) (AJ302566) (AJ302569) (AJ302570)	OR *Homo sapiens*	312	144/298 (48%)	191/298 (63%)	3e-69
Gi|12231029|sp|Q15062|O2H3_HUMAN	OR 2H3 *Homo sapiens*	316	143/294 (48%)	185/294 (62%)	3e-68
Gi|9798920|gb|AAF98752.1|AF211940_1 (AF211940)	OR *Homo sapiens*	303	142/287 (49%)	183/287 (63%)	5e-68

Table 3F. Information for the ClustalW proteins:

1) Novel GPCR3 (SEQ ID NO:10)
2) gi|5051404|emb|CAB45012.1| 573K1.15 (mm17M1-6 (novel 7 transmembrane receptor (rhodopsin family) (olfactory receptor LIKE) protein)) *Mus musculus* (SEQ ID NO:53)
3) gi|12054359|emb|CAC20487.1| olfactory receptor *Homo sapiens* (SEQ ID NO:54)
4) gi|12054355|emb|CAC20485.1| olfactory receptor *Homo sapiens* (SEQ ID NO:55)
5) gi|12231029|sp|Q15062|O2H3_HUMAN OLFACTORY RECEPTOR 2H3 (OLFACTORY RECEPTOR-LIKE PROTEIN FAT11) (SEQ ID NO:56)
6) gi|9798920|gb|AAF98752.1|AF211940_1 olfactory receptor *Homo sapiens* (SEQ ID NO:57)



DOMAIN results for GPCR3 were collected from the Conserved Domain Database (CDD) with Reverse Position Specific BLAST. This BLAST samples domains found in the Smart and Pfam collections. The 7tm_1, a seven transmembrane receptor (rhodopsin family), was shown to have two segments with significant homology to GPCR3. The region from amino acid residue 40 through 226 aligns with amino acids 1 through 167 of the "seven transmembrane receptor (rhodopsin family) fragment" domain(SEQ ID NO:45, E=1e-15), and GPCR3 amino acids 219-290 aligned with residues 305-377 of the 7tm_1 entry of the Pfam database (E=.004). This indicates that the GPCR3 sequence has properties similar to those of other proteins known to contain this domain as well as to the 7tm_1 domain itself.

The nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in various GPCR-related pathological disorders and/or OR-related pathological disorders, described further below. For example, a cDNA encoding the olfactory receptor -like protein may be useful in gene therapy, and the olfactory receptor -like protein may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from neoplasm, adenocarcinoma, lymphoma, prostate cancer, uterus cancer, immune response, AIDS, asthma, Crohn's disease, multiple sclerosis, and Albright Hereditary Ostoeodystrophy. Other GPCR-related diseases and disorders are contemplated.

The novel nucleic acid encoding the GPCR-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-GPCRX Antibodies" section below. For example, the disclosed GPCR3 protein has multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated GPCR3 epitope is from about amino acids 5 to 15. In another embodiment, a GPCR3 epitope is from about amino acids 90 to 100. In additional embodiments, GPCR3 epitopes are from amino acids 175 to 200, 225-250, 255-275 and from amino acids 290 to 325. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

GPCR4

GPCR4 includes a family of three nucleic acids disclosed below. The disclosed nucleic acids encode a GPCR-like protein.

GPCR4a

The disclosed GPCR4a is encoded by three different nucleic acids, GPCR4a1 (dj408b20_C) GPCR4a2 (dj408b20_C_da1), and GPCR4a3 (CG55358-03). A first nucleic acid, dj408b20_C (GPCR4a1), is 947 nucleotides long (SEQ ID NO:11). An open reading frame was identified beginning with an ATG initiation codon at nucleotides 3-5 and ending with a TGA codon at nucleotides 939-941. Putative untranslated regions upstream from the initiation codon and downstream from the termination codon are underlined in Table 4A, and the start and stop

codons are in bold letters. The encoded protein having 312 amino acid residues is presented using the one-letter code in Table 4B (SEQ ID NO:12).

Table 4A. GPCR4a1 Nucleotide Sequence (SEQ ID NO:11).

GT**ATG**GAAAACGATAATACAAGTTCTTTCGAAGGCTTCATCCTGGTGGGCTTCTCTGATCGTCCCCACCT
AGAGCTGATCGTCTTTGTGGTTGTCCTCATCTTTTATCTGCTGACTCTTCTTGGCAACATGACCATTGTC
TTGCTTTCAGCTCTGGATTCCCGGCTGCACACACCAATGTATTTCTTTTTGGCAAACCTCTCATTCCTGG
ACATGTGTTTCACCACAGGTTCCATCCCTCAGATGCTCTACAACCTTTGGGGTCCAGATAAGACCATCAG
CTATGTGGGTTGTGCCATCCAGCTGTACTTTGTCCTGGCCCTGGGAGGGGTGGAGTGTGTCCTCCTGGCT
GTCATGGCATATGACCGCTATGCTGCAGTCTGCAAACCCCTGCACTACACCATCATCATGCACCCACGTC
TCTGTGGACAGCTGGCTTCAGTGGCATGGCTGAGTGGCTTTGGCAATTCTCTCATAATGGCACCCCAGAC
ATTGATGCTACCCCGCTGTGGGCACAGACGAGTTGACCACTTTCTCTGTGAGATGCCAGCACTAATTGGT
ATGGCCTGTGTAGACACCATGATGCTTGAGGCACTGGCTTTTGCCCTGGCAATCTTTATCATCCTGGCAC
CACTCATCCTCATTCTCATTTCTTATGGTTACGTTGGAGGAACAGTGCTTAGGATCAAGTCAGCTGCTGG
GCGAAAGAAAGCCTTCAACACTTGCAGCTCGCATCTAATTGTTGTCTCTCTCTTCTATGGTACAATCATA
TACATGTACCTCCAGCCAGCAAATACTTATTCCCAGGACCAGGGCAAGTTTCTTACCCTTTTCTACACAA
TTGTCACTCCCAGTGTTAACCCCCTGATCTATACACTAAGAAACAAAGATGTTAAAGAGGCCATGAAGAA
GGTGCTAGGGAAGGGGAGTGCAGAAATA**TAG**TAAGGG

The disclosed nucleic acid GPCR4a1 sequence has 624 of 931 bases (67%) identical (with 624/931 positives, 67%) to a 939 bp *Homo sapiens* olfactory receptor-like protein (OR2C1) gene (GENBANK-ID: AF098664) ($E = 3.8e^{-72}$). In a search of sequence databases, partial matches were also identified, *e.g.*, the minus strand of nucleotides 719-947 had 229 of 229 bases (100%) identical to a 1320 bp *synthetic* GPCR mRNA (PATENT-ID: T72050), the sequence around marker 2B8 in HH region of chromosome 6p2.1, and the same region of GPCR4a1 had 229 of 229 bases (100%) identical to a *synthetic* GPCR mRNA (PATENT-ID: T72050), also sequence around marker 2B8 in HH region of chromosome 6p2.1 (E value in both cases is 9.6e-47).

The GPCR4a polypeptide (SEQ ID NO:12) encoded by SEQ ID NO:11 is presented using the one-letter amino acid code in Table 4B. The Psort profile for GPCR4a predicts that this sequence has a signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.6000. The most likely cleavage site for a GPCR4a peptide is between amino acids 41 and 42, *i.e.*, at the slash in the amino acid sequence LLG-NM, based on the SignalP result.

Table 4B. GPCR4a protein sequence (SEQ ID NO:12)

MENDNTSSFEGFILVGFSDRPHLELIVFVVVLIFYLLTLLG/NMTIVLLSALDSRLHTPMYFFLANLSF
LDMCFTTGSIPQMLYNLWGPDKTISYVGCAIQLYFVLALGGVECVLLAVMAYDRYAAVCKPLHYTIIMH
PRLCGQLASVAWLSGFGNSLIMAPQTLMLPRCGHRRVDHFLCEMPALIGMACVDTMMLEALAFALAIFI
ILAPLILILISYGYVGGTVLRIKSAAGRKKAFNTCSSHLIVVSLFYGTIIYMYLQPANTYSQDQGKFLT
LFYTIVTPSVNPLIYTLRNKDVKEAMKKVLGKGSAEI

The predicted GPCR4a1 sequence, above, was subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence

available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such suitable sequences were then used as the forward and reverse primers in a PCR amplification based on a wide range of cDNA libraries. The resulting amplicon was gel purified, cloned and sequenced to high redundancy, as described in the Examples.

The cloned sequence is disclosed as an alternative embodiment of GPCR4a2 (SEQ ID NO:13), referred to herein as the GPCR4a2 and reported in Table 4C. This 945 nucleotide sequence (GPCR4a2) is alternatively referred to herein as dj408b20_C_da1. This nucleic acid is two nucleotides shorter than GPCR4a1 at the 5' UTR: However, CPCR4a2 encodes the same 312 amino acid protein (GPCR4a, SEQ ID NO:12).

Table 4C. GPCR4a2 Nucleotide Sequence (SEQ ID NO:13)

```
GTATGGAAAACGATAATACAAGTTCTTTCGAAGGCTTCATCCTGGTGGGCTTCTCTGATCGTCCCCACCTAGAGCTGATC
GTCTTTGTGGTTGTCCTCATCTTTTATCTGCTGACTCTTCTTGGCAACATGACCATTGTCTTGCTTTCAGCTCTGGATTC
CCGGCTGCACACACCAATGTATTTCTTTTTGGCAAACCTCTCATTCCTGGACATGTGTTTCACCACAGGTTCCATCCCTC
AGATGCTCTACAACCTTTGGGGTCCAGATAAGACCATCAGCTATGTGGGTTGTGCCATCCAGCTGTACTTTGTCCTGGCC
CTGGGAGGGGTGGAGTGTGTCCTCCTGGCTGTCATGGCATATGACCGCTATGCTGCAGTCTGCAAACCCCTGCACTACAC
CATCATCATGCACCCACGTCTCTGTGGACAGCTGGCTTCAGTGGCATGGCTGAGTGGCTTTGGCAATTCTCTCATAATGG
CACCCCAGACATTGATGCTACCCCGCTGTGGGCACAGACGAGTTGACCACTTTCTCTGTGAGATGCCAGCACTAATTGGT
ATGGCCTGTGTAGACACCATGATGCTTGAGGCACTGGCTTTTGCCCTGGCAATCTTTATCATCCTGGCACCACTCATCCT
CATTCTCATTTCTTATGGTTACGTTGGAGGAACAGTGCTTAGGATCAAGTCAGCTGCTGGGCGAAAGAAAGCCTTCAACA
CTTGCAGCTCGCATCTAATTGTTGTCTCTCTCTTCTATGGTACAATCATATACATGTACCTCCAGCCAGCAAATACTTAT
TCCCAGGACCAGGGCAAGTTTCTTACCCTTTTCTACACAATTGTCACTCCCAGTGTTAACCCCCTGATCTATACACTAAG
```

The full amino acid sequence of the disclosed GPCR4a polypeptide has 197 of 305 amino acid residues (64%) identical to, and 242 of 305 residues (79%) positive with, the 320 amino acid residue protein from *Homo sapiens* novel 7 transmembrane receptor (rhodopsin family, olfactory receptor-like protein HS6M1-15, ptnr:SPTREMBL-ACC:Q9Y3N9 DJ88J8.1, E = $5.0e^{-108}$).

BLASTP (Non-Redundant Composite database) analysis of the best hits for alignments with GPCR4a are listed in Table 4D. BLASTX analysis was also performed to determine which proteins have significant identity with GPCR4a, as shown in Table 4E.

Table 4D. BLASTP results for GPCR4a					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
ACC:Q9Y3N9 DJ88J8.1	Novel 7 Transmembrane Receptor (Rhodopsin Family, OR-Like) Hs6m1-15 *Homo Sapiens*	320	197/305 (64%)	242/305 (79%)	9.4e-109
ACC:P23275	OLFACTORY RECEPTOR 15 (OR3) *Mus musculus*	312	190/308 (61%)	241/308 (78%)	5.5e-104
ACC:O76001 DJ80I19.7	OR-LIKE PROTEIN (HS6M1-3) *Homo sapiens*	311	193/302 (63%)	244/302 (80%)	1.5e-103

Table 4E. BLASTX results for GPCR4a				
Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob P(N)	N
Ptnr:SPTREMBL-ACC:Q9Y3N9 DJ88J8.1 (NOVEL 7 TRANSMEMBRA...	+3	1076	5.6e-108	1
Ptnr:SWISSPROT-ACC:P23275 OLFACTORY RECEPTOR 15 (OR3) ...	+3	1031	3.3e-103	1
Ptnr:SPTREMBL-ACC:O76001 DJ80I19.7 (OLFACTORY RECEPTOR...	+3	1027	8.7e-103	1
Ptnr:SPTREMBL-ACC:O95371 OLFACTORY RECEPTOR-LIKE PROTE...	+3	992	4.5e-99	1
Ptnr:SPTREMBL-ACC:O95918 DJ271M21.2 (HS6M1-12 (7 TRANS...	+3	988	1.2e-98	1
Ptnr:SPTREMBL-ACC:Q9WV11 573K1.8 (MM17M1-2 (NOVEL 7 TR...	+3	984	3.2e-98	1
Ptnr:SPTREMBL-ACC:Q9WV14 573K1.2 (MM17M1-3 (NOVEL 7 TR...	+3	981	6.6e-98	1

Possible SNPs found for GPCR4a2 are listed in Table4F.

Table 4F: SNPs		
Base Position	**Base Before**	**Base After**
44	T	C(20
147	T	C(2)
220	T	C(2)
271	T	C(3)
432	A	G(2)
452	C	T(2)
493	A	G(2)
771	T	C(3)

GPCR4b

The target sequence identified as dj408b20_C (GPCR4a1) was again subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. The cDNA coding for the sequence was cloned by polymerase chain reaction (PCR) using the following primers: ATACAAGTTCTTTCGAAGGCTTCATCC (SEQ ID NO:14) and CCCTTACTATATTTCTGCACTCCCCTT (SEQ ID NO:15) on pool 1 of human

cDNAs containing the following: Adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus.

Primers were designed based on in silico predictions for the full length or part (one or more exons) of the DNA/protein sequence of the invention or by translated homology of the predicted exons to closely related human sequences or to sequences from other species. Usually multiple clones were sequenced to derive the sequence which was then assembled similar to the SeqCalling process. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. The PCR product derived by exon linking was cloned into the pCR2.1 vector from Invitrogen. The bacterial clone 115843::DJ408B20_C.698322.D10 has an insert covering the entire open reading frame cloned into the pCR2.1 vector from Invitrogen.

In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on *in silico* predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences from other species. Typically, the resulting amplicons were gel purified, cloned and sequenced to high redundancy as described in the examples. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate.

These procedures provided a third nucleic acid encoding a GPCR4 protein. The nucleic acid is referred to as GPCR4a3 or CG55358-03. This nucleic acid is 932 nucleotides long (SEQ ID NO:16, Table 4G) and is 16 nucleotides shorter in the 5' UTR than GPCR2a1. An open reading frame of the mature protein was identified beginning with an ACA codon which codes for the amino acid threonine at nucleotides 3-5 and ending with a TAG codon at nucleotides 924-926. Putative untranslated regions, if any, are found upstream from the initiation codon and downstream from the termination codon. One silent base substitution is present: C767 is T752 in GPCR4a3. GPCR4b, the protein encoded by GPCR4a3 is identical to GPCR4a, except for the 5 amino acid deletion at the N terminus, as shown in Table 4H (SEQ ID NO:17).

The GPCR Olfactory Receptor disclosed herein is expressed in at least the following tissues: Apical microvilli of the retinal pigment epithelium, arterial (aortic), basal forebrain, brain, Burkitt lymphoma cell lines, corpus callosum, cardiac (atria and ventricle), caudate nucleus, CNS and peripheral tissue, cerebellum, cerebral cortex, colon, cortical neurogenic cells, endothelial (coronary artery and umbilical vein) cells, palate epithelia, eye, neonatal eye, frontal cortex, fetal hematopoietic cells, heart, hippocampus, hypothalamus, leukocytes, liver, fetal liver, lung, lung lymphoma cell lines, fetal lymphoid tissue, adult lymphoid tissue, Those that express MHC II and III nervous, medulla, subthalamic nucleus, ovary, pancreas, pituitary, placenta, pons, prostate, putamen, serum, skeletal muscle, small intestine, smooth muscle (coronary artery in aortic) spinal cord, spleen, stomach, taste receptor cells of the tongue, testis, thalamus, and thymus tissue. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

In a search of sequence databases, it was found, for example, that the disclosed GPCR4a3 nucleic acid sequence has 614 of 913 bases (67%) identical to a gb:GENBANK-ID:AF098664|acc:AF098664.1 mRNA from Homo sapiens (Homo sapiens olfactory receptor-like protein (OR2C1) gene, complete cds, E= 2.1e-71).

Table 4G. GPCR4a3 Nucleotide Sequence (SEQ ID NO:16)

ATACAAGTTCTTTCGAAGGCTTCATCCTGGTGGGCTTCTCTGATCGTCCCCACCTAGAGC	60
TGATCGTCTTTGTGGTTGTCCTCATCTTTTATCTGCTGACTCTTCTTGGCAACATGACCA	120
TTGTCTTGCTTTCAGCTCTGGATTCCCGGCTGCACACACCAATGTATTTCTTTTTGGCAA	180
ACCTCTCATTCCTGGACATGTGTTTCACCACAGGTTCCATCCCTCAGATGCTCTACAACC	240
TTTGGGGTCCAGATAAGACCATCAGCTATGTGGGTTGTGCCATCCAGCTGTACTTTGTCC	300
TGGCCCTGGGAGGGGTGGAGTGTGTCCTCCTGGCTGTCATGGCATATGACCGCTATGCTG	360
CAGTCTGCAAACCCCTGCACTACACCATCATCATGCACCCACGTCTCTGTGGACAGCTGG	420
CTTCAGTGGCATGGCTGAGTGGCTTTGGCAATTCTCTCATAATGGCACCCCAGACATTGA	480
TGCTACCCCGCTGTGGGCACAGACGAGTTGACCACTTTCTCTGTGAGATGCCAGCACTAA	540
TTGGTATGGCCTGTGTAGACACCATGATGCTTGAGGCACTGGCTTTTGCCCTGGCAATCT	600
TTATCATCCTGGCACCACTCATCCTCATTCTCATTTCTTATGGTTACGTTGGAGGAACAG	660
TGCTTAGGATCAAGTCAGCTGCTGGGCGAAAGAAAGCCTTCAACACTTGCAGCTCGCATC	720
TAATTGTTGTCTCTCTCTTCTATGGTACAATTATATACATGTACCTCCAGCCAGCAAATA	780
CTTATTCCCAGGACCAGGGCAAGTTTCTTACCCTTTTCTACACAATTGTCACTCCCAGTG	840
TTAACCCCCTGATCTATACACTAAGAAACAAAGATGTTAAAGAGGCCATGAAGAAGGTGC	900
TAGGGAAGGGGAGTGCAGAAATATAGTAAGGG 932	

The SignalP, Psort and/or Hydropathy profile for the disclosed GPCR4b Olfactory Receptor-like protein predicts that this sequence has a signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.6000. The SignalP shows a signal sequence is coded for in the first 36 amino acids with a cleavage site at between amino acids 36 and 37, as indicated by a slash between LLG/NM in Table 4H. This is typical of this type of membrane protein.

Table 4H. GPCR4b Amino Acid Sequence (SEQ ID NO:17)	
TSSFEGFILVGFSDRPHLELIVFVVVLIFYLLTLLG/NMTIVLLSALDSRLHTPMYFFLAN	60
LSFLDMCFTTGSIPQMLYNLWGPDKTISYVGCAIQLYFVLALGGVECVLLAVMAYDRYAA	120
VCKPLHYTIIMHPRLCGQLASVAWLSGFGNSLIMAPQTLMLPRCGHRRVDHFLCEMPALI	180
GMACVDTMMLEALAFALAIFIILAPLILILISYGYVGGTVLRIKSAAGRKKAFNTCSSHL	240
IVVSLFYGTIIYMYLQPANTYSQDQGKFLTLFYTIVTPSVNPLIYTLRNKDVKEAMKKVL	300
GKGSAEI	307

The full amino acid sequence of the disclosed GPCR4b protein of the invention was found to have 195 of 299 amino acid residues (65%) identical to, and 239 of 299 amino acid residues (79%) similar to, the 320 amino acid residue ptnr:SPTREMBL-ACC:Q9Y3N9 protein from Homo sapiens (DJ88J8.1, NOVEL 7 TRANSMEMBRANE RECEPTOR (RHODOPSIN FAMILY) (OLFACTORY RECEPTOR LIKE) PROTEIN) (HS6M1-15), E= 1.6e-107).

In following positions, one or more consensus positions (Cons. Pos.) of the GPCR4a3 nucleotide sequence have been identified as SNPs. "Depth" represents the number of clones covering the region of the SNP. The Putative Allele Frequency (Putative Allele Freq.) is the fraction of all the clones containing the SNP. A dash ("-"), when shown, means that a base is not present. The sign ">" means "is changed to": Cons. Pos.: 422 Depth: 18 Change: T > C, Putative Allele Freq.: 0.333; Cons. Pos.: 546 Depth: 15 Change: T > C, Putative Allele Freq.: 0.133; Cons. Pos.: 753 Depth: 8 Change: C > T, Putative Allele Freq.: 0.250.

Unless specifically addressed as GPCR4a or GPCR4b, any reference to GPCR4 is assumed to encompass all variants. Residue differences between any GPCRX variant sequences herein are written to show the residue in the "a" variant and the residue position with respect to the "a" variant. In all following sequence alignments, the GPCR4a protein sequence was used.

BLAST results include sequences from the Patp database, which is a proprietary database that contains sequences published in patents and patent publications. The Patp results include those listed in Table 4I.

Table 4I. Patp alignments of GPCR4a

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob. P(N)
patp:B43266 Human ORFX ORF3030 polypeptide sequence SE...	+3	930	1.3e-92
patp:B42796 Human ORFX ORF2560 polypeptide sequence SE..	+3	919	1.8e-91
patp:Y54326 Amino acid sequence of marmot olfactory re...	+3	879	3.2e-87
patp:Y83390 Olfactory receptor protein OLF-5 - H. sapiens...	+3	738	2.8e-72
patp:W75960 Human olfactory OLRCC15 receptor - H. sapiens...	+3	712	1.6e-69
patp:Y83389 Olfactory receptor protein OLF-4 - H. sapiens...	+3	690	3.4e-67

The disclosed GPCR4 protein (SEQ ID NO:12) also has good identity with a number of olfactory receptor proteins. The identity information used for ClustalW analysis is presented in Table 4J.

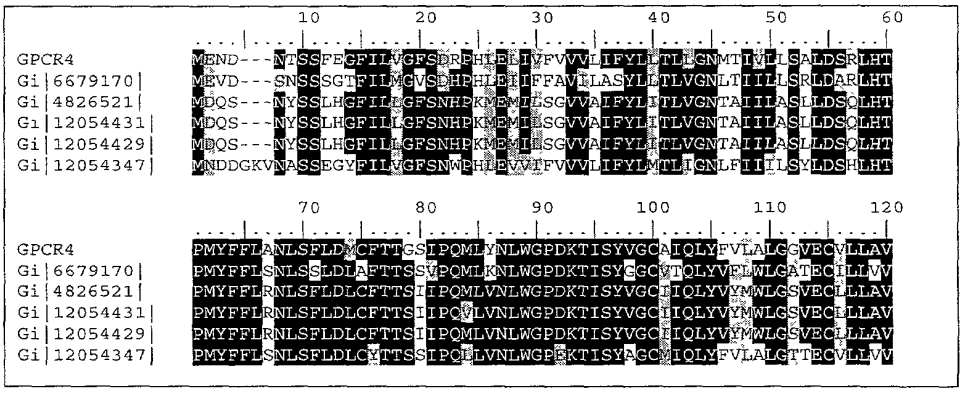
Table 4J. BLAST results for GPCR4

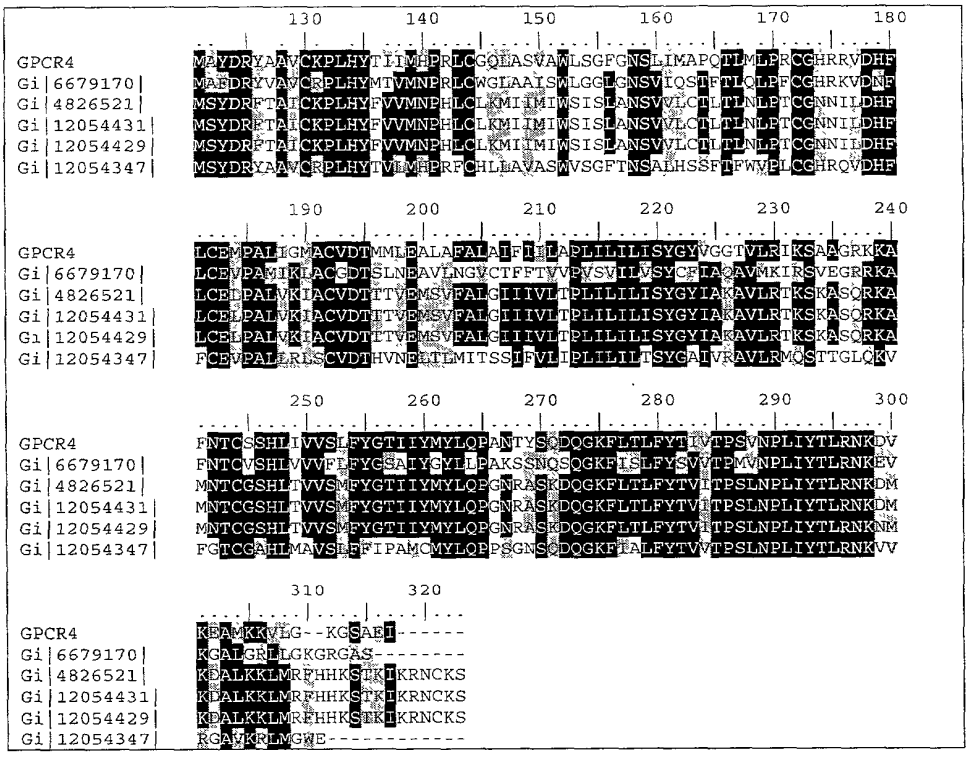
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
Gi|6679170|ref|NP_03 2788.1|	OR 15 (OR3) *Mus musculus*	312	174/308 (56%)	216/308 (69%)	1e-93
Gi|4826521|emb|CAB42 853.1| (AL035402, dJ88J8.1) (AJ302594-99) (AJ302600-01)	Novel 7 tm receptor protein (rhodopsin fam., OR-like) (hs6M1-15) *Homo sapiens*	320	174/305 (57%)	213/305 (69%)	2e-93
Gi|12054431|emb|CAC2 0523.1| (AJ302603)	OR *Homo sapiens*	320	173/305 (56%)	213/305 (69%)	3e-93
Gi|12054429|emb|CAC2 0522.1| (AJ302602)	OR *Homo sapiens*	320	173/305 (56%)	213/305 (69%)	7e-93
Gi|12054347|emb|CAC2 0478.1| (AJ302558)	OR *Homo sapiens*	311	170/302 (56%)	211/302 (69%)	6e-90

This information is presented graphically in the multiple sequence alignment given in Table 4K (with GPCR4 being shown on line 1) as a ClustalW analysis comparing GPCR4 with related OR sequences.

Table 4K Information for the ClustalW proteins:

1) GPCR4 (SEQ ID NO:12)
2) gi|6679170|ref|NP_032788.1| olfactory receptor 15 Mus musculus (SEQ ID NO:58)
3) gi|4826521|emb|CAB42853.1| dJ88J8.1 (novel 7 transmembrane receptor (rhodopsin family) (OR like) protein) (hs6M1-15)) Homo sapiens (SEQ ID NO:59)
4) gi|12054431|emb|CAC20523.1| olfactory receptor Homo sapiens (SEQ ID NO:60)
5) gi|12054429|emb|CAC20522.1| olfactory receptor Homo sapiens (SEQ ID NO:61)
6) gi|12054347|emb|CAC20478.1| olfactory receptor Homo sapiens (SEQ ID NO:62)





DOMAIN results for GPCR4 were collected from the Conserved Domain Database (CDD) with Reverse Position Specific BLAST. This BLAST samples domains found in the Smart and Pfam collections. Two regions of GPCR4 have identity to the 377 amino acid 7TM domain, as described above. The 7tm_1, a seven transmembrane receptor (rhodopsin family), SEQ ID NO:45, above, was shown to have homology to GPCR4. Residues 1-120 of 7TM align with residues 41-159 of GPCR4 (E= 7e-22, shown in Table 2L) and residues 224-290 of GPCR4 have identity with residues 310-377 of 7TM (E=2e-04).

Table 4L: Domain Alignment between GPCR4 and 7TM.

GPCR4	GNMTLVLLSALDSRLHTPMYFFLANLSFLDMCFTTGSIPQMLVNLWGPDKTISYVGCAIQ
TM7	GNVLVCMAVSREKALQTTNYLIVSLAVADLLVATLVMPWVVYLEVVGEWKFSRIHCDIF
GPCR	LYFVTLALGGVECVLLAVMAYDRYAAVCKPLHYTIIM-HPRLCGQLASVAWLSGFGNSLIM
TM7	VTLDVMMCTASIINLCAISIDRYTAVAMPMLYNTRYSSKRRVTVMIAIVWVLSFTISCPM

The nucleic acids and proteins of GPCR4 are useful in potential therapeutic applications implicated in various GPCR-related pathological disorders and/or OR-related pathological disorders, described further below. For example, a cDNA encoding the olfactory receptor-like protein may be useful in gene therapy, and the olfactory receptor-like protein may be useful when administered to a subject in need thereof. The protein similarity information, expression

pattern, and map location for the Olfactory Receptor-like protein and nucleic acid disclosed herein suggest that this Olfactory Receptor may have important structural and/or physiological functions characteristic of the Olfactory Receptor family. Therefore, the nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), and (v) a composition promoting tissue regeneration in vitro and in vivo (vi) biological defense weapon.

The nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications implicated in various diseases and disorders described below and/or other pathologies. For example, the compositions of the present invention will have efficacy for treatment of patients suffering from: developmental diseases, MHCII and III diseases (immune diseases), taste and scent detectability Disorders, Burkitt's lymphoma, corticoneurogenic disease, signal transduction pathway disorders, retinal diseases including those involving photoreception, cell growth rate disorders; cell shape disorders, feeding disorders; control of feeding; potential obesity due to over-eating; potential disorders due to starvation (lack of appetite), noninsulin-dependent diabetes mellitus (NIDDM1), bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease; multiple sclerosis; and treatment of Albright Hereditary Ostoeodystrophy, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy, and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation, dentatorubro-pallidoluysian atrophy (DRPLA) hypophosphatemic rickets, autosomal dominant (2) acrocallosal syndrome and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome and/or other pathologies and disorders of the like. The polypeptides can be used as immunogens to produce antibodies specific for the invention, and as vaccines. They can also be used to screen for potential agonist and antagonist compounds. For example, a cDNA encoding the OR -like protein may be useful in gene therapy, and the OR-like protein may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present

invention will have efficacy for treatment of patients suffering from bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease; multiple sclerosis; and treatment of Albright Hereditary Ostoeodystrophy, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy, and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome. Other GPCR-4 diseases and disorders are contemplated.

The novel nucleic acid encoding OR-like protein, and the OR-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods and other diseases, disorders and conditions of the like. These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-GPCRX Antibodies" section below. For example, the disclosed GPCR4 protein has multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated GPCR4 epitope is from about amino acids 5 to 20. In another embodiment, a GPCR4 epitope is from about amino acids 75 to 90. In additional embodiments, GPCR4 epitopes are from amino acids 120 to 130, 160 to 180, 225-245, 260-280 and from amino acids 290 to 320. These novel proteins can also be used to develop assay system for functional analysis.

GPCR5

GPCR5 includes a family of three similar nucleic acids and three similar proteins disclosed below. The disclosed nucleic acids encode GPCR, OR-like proteins.

GPCR5a

The disclosed novel GPCR5a nucleic acid of 1003 nucleotides (also referred to as 115-a-12-A) is shown in Table 5A. An ORF begins with an ATG initiation codon at nucleotides 6-8 and ends with a TAA codon at nucleotides 999-1001. A putative untranslated region upstream

from the initiation codon and downstream from the termination codon is underlined in Table 5A, and the start and stop codons are in bold letters.

Table 5A. GPCR5a Nucleotide Sequence (SEQ ID NO:18)

AGACA**ATG**AGTCCTGATGGGAACCACAGTAGTGATCCAACAGAGTTCGTCCTGGCAGGGCTCCCAAATCT
CAACAGCGCAAGAGTGGAATTATTTTCTGTGTTTCTTCTTGTCTATCTCCTGAATCTGACAGGCAATGTG
TTGATTGTGGGGGTGGTAAGGGCTGATACTCGACTACAGACCCCTATGTACTTCTTTCTGGGTAACCTGT
CCTGCCTAGAGATACTGCTCACTTCTGTCATCATTCCAAAGATGCTGAGCAATTTCCTCTCAAGGCAACA
CACTATTTCCTTTGCTGCATGTATCACCCAATTCTATTTCTACTTCTTTCTCGGGGCCTCCGAGTTCTTA
CTGTTGGCTGTCATGTCTGCGGATCGCTACCTGGCCATCTGTCATCCTCTGCGCTACCCCTTGCTCATGA
GTGGGGCTGTGTGCTTTCGTGTGGCCTTGGCCTGCTGGGTGGGGGGACTCGTCCCTGTGCTTGGTCCCAC
AGTGGCTGTGGCCTTGCTTCCTTTCTGTAAGCAGGGTGCTGTGGTACAGCACTTCTTCTGCGACAGTGGC
CCACTGCTCCGCCTGGCTTGCACCAACACCAAGAAGCTGGAGGAGACTGACTTTGTCCTGGCCTCCCTCG
TCATTGTATCTTCCTTGCTGATCACTGCTGTGTCCTACGGCCTCATTGTGCTGGCAGTCCTGAGCATCCC
CTCTGCTTCAGGCCGTCAGAAGGCCTTCTCTACCTGTACCTCCCACTTGATAGTGGTGACCCTCTTCTAT
GGAAGTGCCATTTTTCTCTATGTGCGGCCATCGCAGAGTGGTTCTGTGGACACTAACTGGGCAGTGACAG
TAATAACGACATTTGTGACACCACTGTTGAATCCATTCATCTATGCCTTACGTAATGAGCAAGTCAAGGA
AGCTTTGAAGGACATGTTTAGGAAGGTAGTGGCAGGCGTTTTAGGGAATCTTTTACTTGATAAATGTCTC
AGTGAGAAAGCAGTAAAG**TAA**AA

The GPCR5a protein encoded by SEQ ID NO:18 has 331 amino acid residues and is presented using the one-letter code in Table 5B. The Psort profile for GPCR5a predicts that this sequence has a signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.6000, it may also localize to the Golgi body. The most likely cleavage site for a peptide is between amino acids 54 and 55, *i.e.*, at the slash in the amino acid sequence VRA-DT (shown as a slash in Table5B) based on the SignalP result.

Table 5B. Encoded GPCR5a protein sequence (SEQ ID NO:19)

MSPDGNHSSDPTEFVLAGLPNLNSARVELFSVFLLVYLLNLTGNVLIVGVVRA/DTRLQTPMYFF
LGNLSCLEILLTSVIIPKMLSNFLSRQHTISFAACITQFYFYFFLGASEFLLLAVMSADRYLAIC
HPLRYPLLMSGAVCFRVALACWVGGLVPVLGPTVAVALLPFCKQGAVVQHFFCDSGPLLRLACTN
TKKLEETDFVLASLVIVSSLLITAVSYGLIVLAVLSIPSASGRQKAFSTCTSHLIVVTLFYGSAI
FLYVRPSQSGSVDTNWAVTVITTFVTPLLNPFIYALRNEQVKEALKDMFRKVVAGVLGNLLLDKC
LSEKAVK

The disclosed nucleic acid sequence for GPCR5 has 604 of 934 bases (64%) identical to and 604 of 934 bases (64%) positive with *Rattus norvegicus* olfactory receptor protein mRNA (936 bp) (GENBANK-ID: RATOLFPROD| acc:M64378) (E= $1.1e^{-45}$).

The full GPCR5 amino acid sequence has 149 of 304 amino acid residues (49 %) identical to, and 201 of 304 residues (66%) positive with, the 313 amino acid residue olfactory receptor from *Mus musculus* (ptnr: SPTREMBL-ACC: Q9Z1V0) (E= $2.6e^{-74}$).

GPCR5b

GPCR5a (115-a-12-A) was subjected to an exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such suitable sequences were then employed as the forward and reverse primers in a PCR amplification based on a wide range of cDNA libraries. The resulting amplicon was gel purified, cloned and sequenced to high redundancy to provide GPCR5b, which is also referred to as 115-a-12-B.

The nucleotide sequence for GPCR5b (1004 bp, SEQ ID NO:20) is presented in Table 5C. The nucleotide sequence differs from GPCR5a by the addition of a T between A5 and A6, and by 6 nucleotide changes (numbered with respect to GPCR5a) C131 >T; T186>C; G472>A; T579>A; A687>T; C799>T.

Table 5C. GPCR5b Nucleotide Sequence (SEQ ID NO:20)

AGACATATGAGTCCTGATGGGAACCACAGTAGTGATCCAACAGAGTTCGTCCTGGCAGGGCTCCCAA
ATCTCAACAGCGCAAGAGTGGAATTATTTTCTGTGTTTCTTCTTGTCTATCTCCTGAATCTGATAGG
CAATGTGTTGATTGTGGGGGTGGTAAGGGCTGATACTCGACTACAGACCCCCATGTACTTCTTTCTG
GGTAACCTGTCCTGCCTAGAGATACTGCTCACTTCTGTCATCATTCCAAAGATGCTGAGCAATTTCC
TCTCAAGGCAACACACTATTTCCTTTGCTGCATGTATCACCCAATTCTATTTCTACTTCTTTCTCGG
GGCCTCCGAGTTCTTACTGTTGGCTGTCATGTCTGCGGATCGCTACCTGGCCATCTGTCATCCTCTG
CGCTACCCCTTGCTCATGAGTGGGGCTGTGTGCTTTCGTGTGGCCTTGGCCTGCTGGGTGGGGGGAC
TCATCCCTGTGCTTGGTCCCACAGTGGCTGTGGCCTTGCTTCCTTTCTGTAAGCAGGGTGCTGTGGT
ACAGCACTTCTTCTGCGACAGTGGCCCACTGCTCCGCCTGGCATGCACCAACACCAAGAAGCTGGAG
GAGACTGACTTTGTCCTGGCCTCCCTCGTCATTGTATCTTCCTTGCTGATCACTGCTGTGTCCTACG
GCCTCATTGTGCTGGCTGTCCTGAGCATCCCCTCTGCTTCAGGCCGTCAGAAGGCCTTCTCTACCTG
TACCTCCCACTTGATAGTGGTGACCCTCTTCTATGGAAGTGCCATTTTTCTCTATGTGCGGTCATCG
CAGAGTGGTTCTGTGGACACTAACTGGGCAGTGACAGTAATAACGACATTTGTGACACCACTGTTGA
ATCCATTCATCTATGCCTTACGTAATGAGCAAGTCAAGGAAGCTTTGAAGGACATGTTTAGGAAGGT
AGTGGCAGGCGTTTTAGGGAATCTTTTACTTGATAAATGTCTCAGTGAGAAAGCAGTAAAGTAAAA

The encoded GPCR5b protein is presented in Table 5D. The disclosed protein is 331 amino acids long and is denoted by SEQ ID NO:21. GPCR5b differs from GPCR5a by 3 amino acid residues: T42>I; V151>I; P265>S. Like GPCR5a, the Psort profile for GPCR5b predicts that this sequence has a signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.6000. The most likely cleavage site for a peptide is between amino acids 54 and 55, *i.e.*, at the slash in the amino acid sequence VRA-DT (shown as a slash in Table5D) based on the SignalP result.

Table 5D. Encoded GPCR5b protein sequence (SEQ ID NO:21)
MSPDGNHSSDPTEFVLAGLPNLNSARVELFSVFLLVYLLNLIGNVLIVGVVRA/DTRLQTPMYFFLGN LSCLEILLTSVIIPKMLSNFLSRQHTISFAACITQFYFYFFLGASEFLLLAVMSADRYLAICHPLRYP LLMSGAVCFRVALACWVGGLIPVLGPTVAVALLPFCKQGAVVQHFFCDSGPLLRLACTNTKKLEETDF VLASLVIVSSLLITAVSYGLIVLAVLSIPSASGRQKAFSTCTSHLIVVTLFYGSAIFLYVRSSQSGSV DTNWAVTVITTFVTPLLNPFIYALRNEQVKEALKDMFRKVVAGVLGNLLLDKCLSEKAVK

BLASTP (Non-Redundant Composite database) analysis of the best hits for alignments with GPCR5b are listed in Table 5E.

Table 5E. BLASTP results for GPCR5b

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
SPTREMBL-ACC:Q9Z1V0	OLFACTORY RECEPTOR C6 *Mus musculus*	313	150/304 (49%)	203/304 (66%)	9.4e-76
SWISSPROT-ACC:P23267	OLFACTORY RECEPTOR-LIKE PROTEIN F6 *Rattus norvegicus*	311	152/301 (50%)	199/301 (66%)	5.2e-75
SPTREMBL-ACC:Q9Y3P5 DJ994E9.5	HS6M1-17 (NOVEL 7 TM (RHODOPSIN FAMILY) (OR LIKE PROTEIN) *Homo sapiens*	306	145/301 (48%)	197/301 (65%)	1.1e-67

A BLASTX was also performed to determine the proteins that have significant identity with GPCR4a. The BLASTX results are shown in Table 5F.

Table 5F. BLASTX results for GPCR5b

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob P(N)	N
Ptnr:SPTREMBL-ACC:Q9Z1V0 OLFACTORY RECEPTOR C6 - Mus m...	+1	764	5.6e-75	1
ptnr:SWISSPROT-ACC:P23267 OLFACTORY RECEPTOR-LIKE PROT...	+1	757	3.1e-74	1
ptnr:SWISSPROT-ACC:P23270 OLFACTORY RECEPTOR-LIKE PROT...	+1	701	2.7e-68	1
ptnr:SPTREMBL-ACC:Q9Y3P5 DJ994E9.5 (HS6M1-17 (NOVEL 7 ...	+1	688	6.3e-67	1
ptnr:SPTREMBL-ACC:O70271 OLFACTORY RECEPTOR-LIKE PROTE...	+1	688	6.3e-67	1
ptnr:SPTREMBL-ACC:O95007 WUGSC:H_DJ0669B10.3 PROTEIN -...	+1	680	4.5e-66	1
ptnr:SPTREMBL-ACC:O13036 CHICK OLFACTORY RECEPTOR 7 - ...	+1	675	1.5e-65	1
ptnr:SPTREMBL-ACC:O95222 OLFACTORY RECEPTOR - Homo sap...	+1	672	3.1e-65	1
ptnr:SPTREMBL-ACC:O70269 OLFACTORY RECEPTOR-LIKE PROTE...	+1	670	5.1e-65	1
ptnr:SPTREMBL-ACC:O57597 CHICK OLFACTORY RECEPTOR 7 - ...	+1	669	6.5e-65	1
ptnr:SPTREMBL-ACC:O70270 OLFACTORY RECEPTOR-LIKE PROTE...	+1	668	8.4e-65	1
ptnr:TREMBLNEW-ACC:AAF65461 OLFACTORY RECEPTOR P2 - Mu...	+1	666	1.4e-64	1
ptnr:SPTREMBL-ACC:Q9WU86 ODORANT RECEPTOR S1 - Mus mus...	+1	665	1.7e-64	1
ptnr:SPTREMBL-ACC:Q90808 OLFACTORY RECEPTOR 4 - Gallus...	+1	665	1.7e-64	1
ptnr:SWISSPROT-ACC:P37071 OLFACTORY RECEPTOR-LIKE PROT...	+1	646	1.8e-62	1

GPCR5c

Another nucleotide sequence resulted when GPCR5a (115-a-12-A) was subjected to an exon linking process to confirm the sequence. PCR primers were designed by starting at the

most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such suitable sequences were then employed as the forward and reverse primers in a PCR amplification based on a wide range of cDNA libraries. The resulting amplicon was gel purified, cloned and sequenced to high redundancy to provide the sequence reported below, which is designated as Accession Number 115_A_12_A_da1, or GPCR5c.

Human tissues providing SeqCalling Fragments of the clone include Pool One: adrenal gland, bone marrow, brain – amygdala, brain – cerebellum, brain – hippocampus, brain - substantia nigra, brain – thalamus, brain – whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma – Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. The tissue origin of the clone is RACE(asm:126603384).

The nucleotide sequence for GPCR5c (1005 bp, SEQ ID NO:22) is presented in Table 5G. The GPCR5c nucleotide sequence differs from GPCR5a by having an extra T at the 5' end, an A at the 3' end, and 6 nucleotide changes: (numbered with respect to GPCR5a) T123>C; C131>T; T186>C; G472>A; T579>A; A687>T.

Table 5G. GPCR5c Nucleotide Sequence (SEQ ID NO:22)

```
TAGACAATGAGTCCTGATGGGAACCACAGTAGTGATCCAACAGAGTTCGTCCTGGCAGGGCTCCCAAATC
TCAACAGCGCAAGAGTGGAATTATTTTCTGTGTTTCTTCTTGTCTATCTCCCGAATCTGATAGGCAATGT
GTTGATTGTGGGGGTGGTAAGGGCTGATACTCGACTACAGACCCCCATGTACTTCTTTCTGGGTAACCTG
TCCTGCCTAGAGATACTGCTCACTTCTGTCATCATTCCAAAGATGCTGAGCAATTTCCTCTCAAGGCAAC
ACACTATTTCCTTTGCTGCATGTATCACCCAATTCTATTTCTACTTCTTTCTCGGGGCCTCCGAGTTCTT
ACTGTTGGCTGTCATGTCTGCGGATCGCTACCTGGCCATCTGTCATCCTCTGCGCTACCCCTTGCTCATG
AGTGGGGCTGTGTGCTTTCGTGTGGCCTTGGCCTGCTGGGTGGGGGGACTCATCCCTGTGCTTGGTCCCA
CAGTGGCTGTGGCCTTGCTTCCTTTCTGTAAGCAGGGTGCTGTGGTACAGCACTTCTTCTGCGACAGTGG
CCCACTGCTCCGCCTGGCATGCACCAACACCAAGAAGCTGGAGGAGACTGACTTTGTCCTGGCCTCCCTC
GTCATTGTATCTTCCTTGCTGATCACTGCTGTGTCCTACGGCCTCATTGTGCTGGCTGTCCTGAGCATCC
CCTCTGCTTCAGGCCGTCAGAAGGCCTTCTCTACCTGTACCTCCCACTTGATAGTGGTGACCCTCTTCTA
TGGAAGTGCCATTTTTCTCTATGTGCGGCCATCGCAGAGTGGTTCTGTGGACACTAACTGGGCAGTGACA
GTAATAACGACATTTGTGACACCACTGTTGAATCCATTCATCTATGCCTTACGTAATGAGCAAGTCAAGG
AAGCTTTGAAGGACATGTTTAGGAAGGTAGTGGCAGGCGTTTTAGGGAATCTTTTACTTGATAAATGTCT
CAGTGAGAAAGCAGTAAAGTAAAAA
```

The coding region of GPCR5c is from nucleotide 7 to 1000, giving the encoded GPCR5c protein, as presented in Table 5H. The disclosed protein is 331 amino acids long and is denoted by SEQ ID NO: 23. GPCR5c differs from GPCR5a by 3 amino acid residues: L39>P; T42>I; and V151>I. Like GPCR5a, the Psort profile for GPCR5c predicts that this sequence has a signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.6000.

The most likely cleavage site for a peptide is between amino acids 54 and 55, *i.e.*, at the slash in the amino acid sequence VRA-DT (shown as a slash in Table5H) based on the SignalP result.

Table 5H. Encoded GPCR5b protein sequence (SEQ ID NO:23)

MSPDGNHSSDPTEFVLAGLPNLNSARVELFSVFLLVYLPNLIGNVLIVGVVRA/DTRLQTPMYFFLGNLSC
LEILLTSVIIPKMLSNFLSRQHTISFAACITQFYFYFFLGASEFLLLAVMSADRYLAICHPLRYPLLMSG
AVCFRVALACWVGGLIPVLGPTVAVALLPFCKQGAVVQHFFCDSGPLLRLACTNTKKLEETDFVLASLVI
VSSLLITAVSYGLIVLAVLSIPSASGRQKAFSTCTSHLIVVTLFYGSAIFLYVRPSQSGSVDTNWAVTVI
TTFVTPLLNPFIYALRNEQVKEALKDMFRKVVAGVLGNLLLDKCLSEKAVK

Patp results include those listed in Table 5I.

Table 5I. Patp alignments of GPCR5

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob. P(N)
patp:Y96680 Murine olfactory receptor ligand-binding...	+3	755	4.4e-74
patp:R27869 Odorant receptor clone F6 - Rattus rattus,...	+3	748	2.4e-73
patp:R27872 Odorant receptor clone I7 - Rattus rattus,...	+3	715	7.6e-70
patp:W21666 Rat spermatid chemoreceptor G-X - Rattus sp....	+3	673	2.2e-65
patp:R48741 G-protein coupled odorant receptor F6 prot...	+3	654	2.2e-63
patp:W02713 G-protein coupled odorant receptor F6 - ...	+3	654	2.2e-63

For example, a BLAST against Y96680, a 313 amino acid olfactory receptor (ligand binding region) from *Mus musculus*, produced 149/304 (49%) identity, and 201/304 (66%) positives (E = 4.4e-74), with long segments of amino acid identity. WO 00/35274.

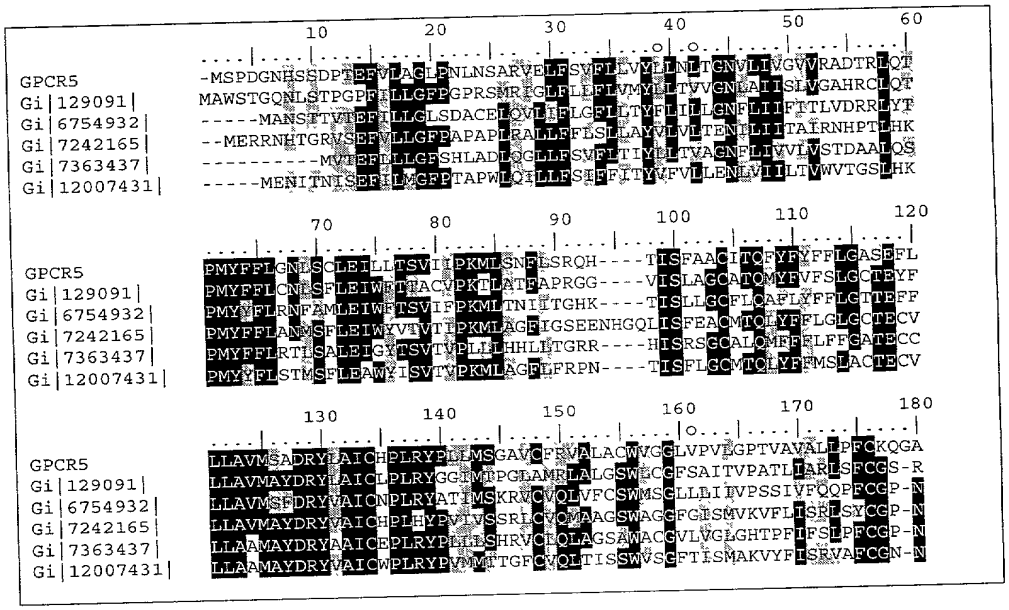
The disclosed GPCR5 protein (SEQ ID NO:19) has good identity with a number of olfactory receptor proteins. The identity information used for ClustalW analysis is presented in Table 5J. Unless specifically addressed as GPCR5a GPCR5b, or GPCR5c, any reference to GPCR5 is assumed to encompass all variants. Residue differences between any GPCRX variant sequences herein are written to show the residue in the "a" variant and the residue position with respect to the "a" variant. GPCR residues in all following sequence alignments that differ between the individual GPCR variants are highlighted with a box and marked with the (o) symbol above the variant residue in all alignments herein. For example, the protein shown in line 1 of Table 5K depicts the sequence for GPCR5a, and the positions where GPCR5b or GPCR5c differs are marked with a (o) symbol and are highlighted with a box. All GPCR5 proteins have significant homology to olfactory receptor (OR) proteins:

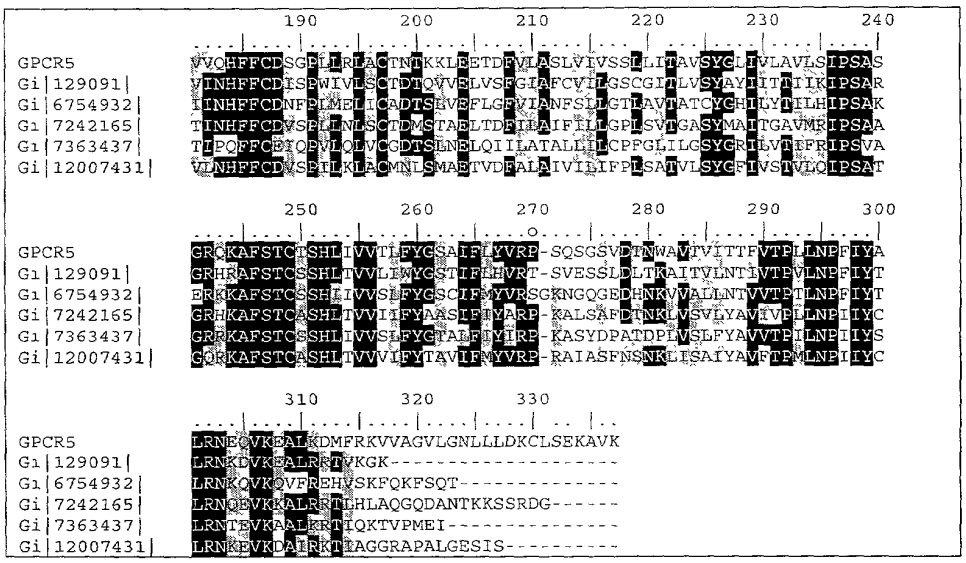
Table 5J. BLAST results for GPCR5					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
Gi|129091|sp|P23267 |OLF6_RAT	OR 15 (OR3) *Rattus norvegicus*	311	144/301 (47%)	189/301 (61%)	7e-67
Gi|6754932|ref|NP_0 35121.1| (AF102523)	OR 49, ORC6 Mus musculus	313	145/305 (47%)	192/305 (62%)	3e-66
Gi|7242165|ref|NP_0 35113.1| (AF106007) (AF321233)	OR 41, ORI7 *Mus musculus*	234	136/304 (44%)	181/304 (58%)	9e-61
Gi|7363437|ref|NP_0 39229.1|	OR, family 10, subfamily C, member 1 *Homo sapiens*	306	139/301 (46%)	186/301 (61%)	2e-59
Gi|12007431|gb|AAG4 5202.1|AF321236_1 (AF321236)	M50 OR *Mus musculus*	316	129/303 (42%)	181/303 (59%)	4e-59

This information is presented graphically in the multiple sequence alignment given in Table 5K (with GPCR5 being shown on line 1) as a ClustalW analysis comparing GPCR5 with related protein sequences.

Table 5K Information for the ClustalW proteins:

1) GPCR5 (SEQ ID NO:19)
2) gi|129091|sp|P23267|OLF6_RAT OLFACTORY RECEPTOR-LIKE PROTEIN F6 (SEQ ID NO:63)
3) gi|6754932|ref|NP_035121.1| olfactory receptor 49 Mus musculus (SEQ ID NO:64)
4) gi|7242165|ref|NP_035113.1| olfactory receptor 41 Mus musculus (SEQ ID NO:65)
5) gi|7363437|ref|NP_039229.1| olfactory receptor, family 10, subfamily C, member 1 Homo sapiens (SEQ ID NO:66)
6) gi|12007431|gb|AAG45202.1|AF321236_1 m50 olfactory receptor Mus musculus (SEQ ID NO:67)



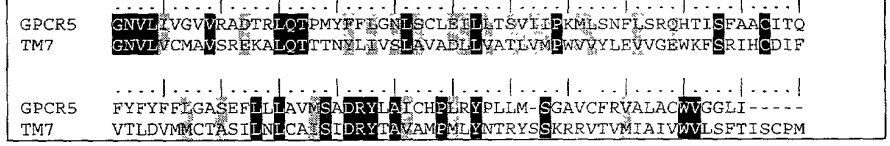


DOMAIN results for GPCR5 were collected from the Conserved Domain Database (CDD) with Reverse Position Specific BLAST. This BLAST samples domains found in the Smart and Pfam collections. The results are listed in Table 5L with the statistics and domain description.

Residues 1-115 of 7tm_1 (SEQ ID NO:45) are aligned with GPCR5 43-156 (E = 1e-20), in Table 5K. Residues 314-377 of 7tm_1 also have identity with residues 231-293 (E=2e-05) of GPCR5.

Table 5L. DOMAIN results for GPCR5

Gnl|Pfam|pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family) (SEQ ID NO: 45) Length = 377
Statistics for GPCR5: Score = 93.6 bits (231), Expect = 1e-20



The nucleic acids and proteins of GPCR5 are useful in potential therapeutic applications implicated in various GPCR-related pathological disorders and/or OR-related pathological disorders, described further below. For example, a cDNA encoding the GPCR (or olfactory-receptor) like protein may be useful in gene therapy, and the receptor -like protein may be useful when administered to a subject in need thereof. The nucleic acids and proteins of the invention are also useful in potential therapeutic applications used in the treatment of infections such as

bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease; multiple sclerosis; and treatment of Albright Hereditary Ostoeodystrophy, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy, and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome and/or other pathologies and disorders. Other GPCR-related diseases and disorders are contemplated.

The polypeptides can be used as immunogens to produce antibodies specific for the invention, and as vaccines. They can also be used to screen for potential agonist and antagonist compounds. For example, a cDNA encoding the GPCR-like protein may be useful in gene therapy, and the GPCR-like protein may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease; multiple sclerosis; and treatment of Albright Hereditary Ostoeodystrophy, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy, and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome and/or other pathologies and disorders. The novel nucleic acid encoding GPCR-like protein, and the GPCR-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-GPCRX Antibodies" section below. For example the disclosed GPCR5 protein has multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated GPCR5 epitope is from about amino acids 5 to 20. In another embodiment, a GPCR5 epitope is from about amino acids 55 to 65. In additional

embodiments, GPCR5 epitopes are from amino acids 175 to 205, 225 to 245, 260-280 and from amino acids 290 to 320. This novel protein also has value in development of powerful assay system for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

GPCR6

The disclosed novel GPCR6 nucleic acid of 948 nucleotides (also referred to as 6-L-19-C) is shown in Table 6A. An open reading begins with an ATG initiation codon at nucleotides 7-9 and ends with a TAG codon at nucleotides 940-942. A putative untranslated region upstream from the initiation codon and downstream from the termination codon are underlined in Table 6A, and the start and stop codons are in bold letters.

Table 6A. GPCR6 Nucleotide Sequence (SEQ ID NO:24)

GGAGAC**ATG**GGCAAGGAAAACTGCACCACTGTGGCTGAGTTCATTCTCCTTGGACTATCAGATGTCCCTG
AGTTGAGAGTCTGCCTCTTCCTGCTGTTCCTTCTCATCTATGGAGTCACGTTGTTAGCCAACCTGGGCAT
GATTGCACTGATTCAGGTCAGCTCTCGGCTCCACACCCCCATGTACTTTTTCCTCAGCCACTTGTCCTCT
GTAGATTTCTGCTACTCCTCAATAATTGTGCCAAAAATGTTGGCTAATATCTTTAACAAGGACAAAGCCA
TCTCCTTCCTAGGGTGCATGGTGCAATTCTACTTGTTTTGCACTTGTGTGGTCACTGAGGTCTTCCTGCT
GGCCGTGATGGCCTATGACCGCTTTGTGGCCATCTGTAACCCTTTGCTATACACAGTCACCATGTCTTGG
AAGGTGCGTGTGGAGCTGGCTTCTTGCTGCTACTTCTGTGGGACGGTGTGTTCTCTGATTCATTTGTGCT
TAGCTCTTAGGATCCCCTTCTATAGATCTAATGTGATTAACCACTTTTTCTGTGATCTACCTCCTGTCTT
AAGTCTTGCTTGCTCTGATATCACTGTGAATGAGACACTGCTGTTCCTGGTGGCCACTTTGAATGAGAGT
GTTACCATCATGATCATCCTCACCTCCTACCTGCTAATTCTCACCACCATCCTGAAGATGGGCTCTGCAG
AGGGCAGGCACAAAGCCTTCTCCACCTGTGCTTCCCACCTCACAGCTATCACTGTCTTCCATGGAACAGT
CCTTTCCATTTATTGCAGGCCCAGTTCAGGCAATAGTGGAGATGCTGACAAAGTGGCCACCGTGTTCTAC
ACAGTCGTGATTCCTATGCTGAACTCTGTGATCTACAGCCTGAGAAATAAAGATGTGAAAGAAGCTCTCA
GAAAAGTGATGGGCTCCAAAATTCACTCC**TAG**GGAAGA

The disclosed nucleic acid sequence has 617 of 915 bases (67%) identical to a *G. gallus* cor4 olfactory receptor 4 DNA (GENBANK-ID: GGCOR4GEN|acc:X94744) (E value = $8.7e\text{-}^{65}$).

The GPCR6 protein encoded by SEQ ID NO:24 has 318 amino acid residues, and is presented using the one-letter code in Table 6B (SEQ ID NO:25). The SignalP, Psort and/or Hydropathy profile for GPCR6 predict that GPCR6 has a signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.6000. The SignalP shows a signal sequence is coded for in the first 41 amino acids, *i.e.*, with a cleavage site at the slash in the sequence TLL-AN, between amino acids 40 and 41. This is typical of this type of membrane protein.

Table 6B. Encoded GPCR6 protein sequence (SEQ ID NO:25).
MGKENCTTVAEFILLGLSDVPELRVCLFLLFLLIYGVTLL/ANLGMIALIQVSSRLHTPMYFFLSHLSSVD FCYSSIIVPKMLANIFNKDKAISFLGCMVQFYLFCTCVVTEVFLLAVMAYDRFVAICNPLLYTVTMSWKV RVELASCCYFCGTVCSLIHLCLALRIPFYRSNVINHFFCDLPPVLSLACSDITVNETLLFLVATLNESVT IMIILTSYLLILTTILKMGSAEGRHKAFSTCASHLTAITVFHGTVLSIYCRPSSGNSGDADKVATVFYTV VIPMLNSVIYSLRNKDVKEALRKVMGSKIHS

The full amino acid sequence of the protein of the invention was found to have 166 of 307 amino acid residues (54%) identical to, and 217 of 307 residues (70%) positive with, the 314 amino acid residue human olfactory receptor-like protein OLF1 (ptnr:SWISSPROT-ACC:Q13606) (E value = $5.8e^{-86}$).

The GPCR6 target sequence identified previously (6_L_19_C) was subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached.

The cDNA coding for the sequence was cloned by polymerase chain reaction (PCR) using the following primers: CACTGTGGCTGAGTTCATTCTCCTT (SEQ ID NO:26) and TCTTCCCTAGGAGTGAATTTTGGAGC (SEQ ID NO:27) on the following pool of human cDNAs: Pool 1 - Adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Primers were designed based on in silico predictions for the full length or part (one or more exons) of the DNA/protein sequence of the invention or by translated homology of the predicted exons to closely related human sequences or to sequences from other species. Usually multiple clones were sequenced to derive the sequence which was then assembled similar to the SeqCalling process. In addition, sequence traces were evaluated manually and edited for corrections if appropriate.

The PCR product derived by exon linking was cloned into the pCR2.1 vector from Invitrogen. The bacterial clone 55446::6_L_19_C.698018.M1 has an insert covering the entire open reading frame cloned into the pCR2.1 vector from Invitrogen.

Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The resulting sequences from all clones were assembled with themselves, with

other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported below, which is designated Accession Number CG50383_01 which does not differ from GPCR6 (6_L_19_C) in amino acid or nucleotide sequence.

The disclosed GPCR6-Olfactory Receptor-like protein is expressed in at least the following tissues: Apical microvilli of the retinal pigment epithelium, arterial (aortic), basal forebrain, brain, Burkitt lymphoma cell lines, corpus callosum, cardiac (atria and ventricle), caudate nucleus, CNS and peripheral tissue, cerebellum, cerebral cortex, colon, cortical neurogenic cells, endothelial (coronary artery and umbilical vein) cells, palate epithelia, eye, neonatal eye, frontal cortex, fetal hematopoietic cells, heart, hippocampus, hypothalamus, leukocytes, liver, fetal liver, lung, lung lymphoma cell lines, fetal lymphoid tissue, adult lymphoid tissue, tissues that express MHC II and III, nervous tissue, medulla, subthalamic nucleus, ovary, pancreas, pituitary, placenta, pons, prostate, putamen, serum, skeletal muscle, small intestine, smooth muscle (coronary artery in aortic) spinal cord, spleen, stomach, taste receptor cells of the tongue, testis, thalamus, and thymus tissue. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

Patp results include those listed in Table 6C.

Table 6C. Patp alignments of GPCR6

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob. P(N)
patp:Y92364 G protein-coupled receptor protein 4 - H. sap	.. +2	668	7.3e-65
patp:Y90872 Human G protein-coupled receptor GTAR14-1 ..	. +2	664	1.9e-64
patp:Y90872 Human G protein-coupled receptor GTAR14-1 ..	. +2	664	1.9e-64
patp:Y54329 Amino acid sequence of marmot olfactory re..	. +2	533	1.5e-50
patp:Y90874 Human G protein-coupled receptor GTAR14-5 ..	. +2	510	4.0e-48

For example, a BLAST against Y923264, a 307 amino acid G-protein coupled receptor from *Homo sapiens*, produced 131/299 (43%) identity, and 195/299 (65%) positives (E = 7.3e-65), with long segments of amino acid identity. WO 00/20590.

The disclosed GPCR6 protein (SEQ ID NO:25) has good identity with a number of olfactory receptor proteins. The identity information used for ClustalW analysis is presented in Table 6D. The GPCR6 protein has significant identity to olfactory receptor (OR) proteins:

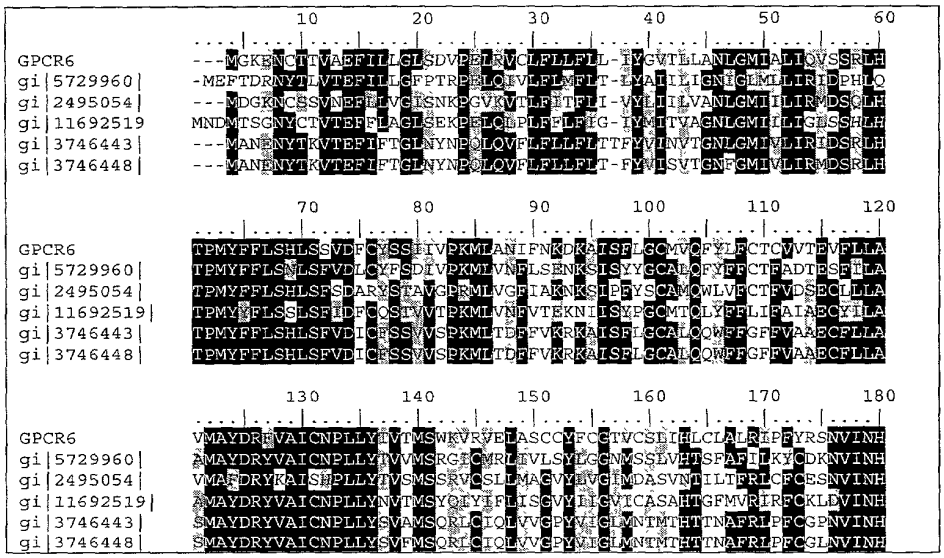
Table 6D. BLAST results for GPCR6

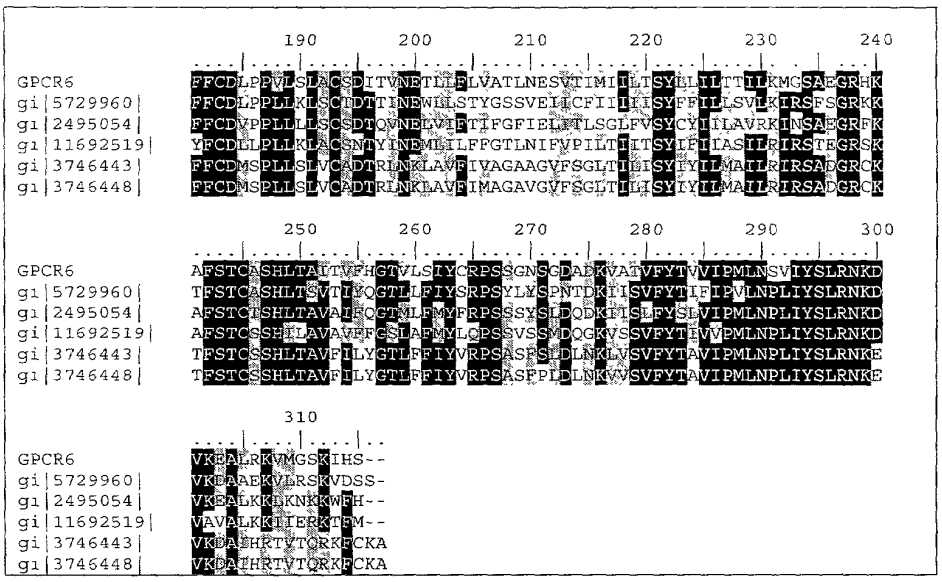
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
Gi|5729960|ref|NP_0 06628.1|	OR fam. 5, subfam. I, mem 1 *Homo sapiens*	314	154/307 (50%)	200/307 (64%)	2e-72
Gi|2495054|sp|Q9515 5|OLF2_CANFA	OR-like prt OLF2 *Canis familiaris*	311	152/308 (49%)	207/308 (66%)	8e-70
Gi|11692519|gb|AAG3 9856.1|AF282271_1 (AF282271)	OR 41, K11 *Mus musculus*	314	150/308 (48%)	199/308 (63%)	8e-69
Gi|3746443|gb|AAC63 969.1| (AF045577)	OR, OR93ch *Pan troglodytes*	314	151/306 (49%)	199/306 (64%)	1e-68
Gi|3746448|gb|AAC63 971.1| (AF045580)	OR OR93Gib *Hylobates lar*	313	149/305 (48%)	198/305 (64%)	3e-68

This information is presented graphically in the multiple sequence alignment given in Table 6E (with GPCR6 being shown on line 1) as a ClustalW analysis comparing GPCR6 with related protein sequences.

Table 6E Information for the ClustalW proteins:

1) GPCR6 (SEQ ID NO:25)
2) gi|5729960|ref|NP_006628.1| OR, family 5, subfamily I, member 1 Homo sapiens (SEQ ID NO:25)
3) gi|2495054|sp|Q95155|OLF2_CANFA OR-LIKE PROTEIN OLF2 (SEQ ID NO:25)
4) gi|11692519|gb|AAG39856.1|AF282271_1 odorant receptor K11 Mus musculus (SEQ ID NO:25)
5) gi|3746443|gb|AAC63969.1| olfactory receptor OR93Ch [Pan troglodytes] (SEQ ID NO:25)
6) gi|3746448|gb|AAC63971.1| olfactory receptor OR93Gib [Hylobates lar] (SEQ ID NO:25)





The presence of identifiable domains in GPCR6 was determined by searches using algorithms such as PROSITE, Blocks, Pfam, ProDomain, Prints and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (http:www.ebi.ac.uk/interpro/).

DOMAIN results for GPCR6 were collected from the Conserved Domain Database (CDD) with Reverse Position Specific BLAST. This BLAST samples domains found in the Smart and Pfam collections. The results are listed in Table 6F with the statistics and domain description. The results indicate that this protein contains the 7tm_1 (InterPro) 7 transmembrane receptor (rhodopsin family) (as defined by Interpro) at residues 42-203, which align with residues 2-158 of the 7TM domain. This indicates that the sequence of GPCR6 has properties similar to those of other proteins known to contain this/these domain(s) and similar to the properties of these domains.

Table 6F. DOMAIN results for GPCR6

gnl|Pfam|pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family) Length = 377 (SEQ ID NO:45).

Score = 86.3 bits (212), Expect = 2e-18

```
GPCR6 -NLGMIALIQVSSRLHTPMYFFLSHLSSVDFCYSSIIVPKMLANIFNKDKAISFLGCMVQ
TM7   GNVLVCMAVSREKALQTTTNYLIVSLAVADLLVATLVMPWVVYLEVVGEWKFSRIHCDIF

GPCR6 FYLFCTCVVTEVFLLAVMAYDRFVAICNPLLYTVTMSWKVRVELAS-CCYFCGTVCSLIH
TM7   VTLDVMMCTASILNLCAISIDRYTAVAMPMLYNTRYSSKRRVTVMIAIVWVLSFTISCPM
```

The similarity information for the GPCR6 protein and nucleic acid disclosed herein suggest that GPCR6 may have important structural and/or physiological functions characteristic of the Olfactory Receptor family and the GPCR family. Therefore, the nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), and (v) a composition promoting tissue regeneration in vitro and in vivo (vi) biological defense weapon. The novel nucleic acid encoding GPCR6, and the GPCR6 protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

The nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications implicated in various diseases and disorders described below and/or other pathologies. For example, the compositions of the present invention will have efficacy for treatment of patients suffering from: developmental diseases, MHCII and III diseases (immune diseases), taste and scent detectability disorders, Burkitt's lymphoma, corticoneurogenic disease, signal transduction pathway disorders, retinal diseases including those involving photoreception, cell growth rate disorders; cell shape disorders, feeding disorders; control of feeding; potential obesity due to over-eating; potential disorders due to starvation (lack of appetite), noninsulin-dependent diabetes mellitus (NIDDM1), bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease; multiple sclerosis; and treatment of Albright hereditary ostoeodystrophy, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy, and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation dentatorubro-pallidoluysian atrophy(DRPLA) hypophosphatemic rickets, autosomal dominant (2) acrocallosal syndrome and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome and/or other pathologies and disorders of the like.

The polypeptides can be used as immunogens to produce antibodies specific for the invention, and as vaccines. They can also be used to screen for potential agonist and antagonist

compounds. For example, a cDNA encoding the OR -like protein may be useful in gene therapy, and the OR-like protein may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to Neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease; multiple sclerosis; and treatment of Albright hereditary ostoeodystrophy, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy, and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome and/or other pathologies and disorders.

The novel nucleic acid encoding OR-like protein, and the OR-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the novel GPCR6 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-GPCRX Antibodies" section below. In one embodiment, a contemplated GPCR6 epitope is from about aa 160 to 170. In another embodiment, a GPCR6 epitope is from about aa 225 to 240. In additional embodiments, GPCR6 epitopes are from aa 255 to 280 and from 280 to 320.

GPCR7

A novel GPCR nucleic acid was identified by TblastN using CuraGen Corporation's sequence file for GPCR probes or homologs, and run against the Genomic Daily Files made available by GenBank. The nucleic acid was further predicted by the program GenScan™, including selection of exons. These were further modified by means of similarities using BLAST searches. The sequences were then manually corrected for apparent inconsistencies, thereby obtaining the sequences encoding the full-length protein. The disclosed novel GPCR7 nucleic acid of 1013 nucleotides (also referred to as dj313i6_D) is shown in Table 7A. An open reading begins with an ATG initiation codon at nucleotides 5-7 and ends with a TAG codon at nucleotides 997-999. A putative untranslated region upstream from the initiation codon and

downstream from the termination codon are underlined in Table 7A, and the start and stop codons are in bold letters.

Table 7A. GPCR7 Nucleotide Sequence (SEQ ID NO:28)

TACA**ATG**GAAAGAGCTAACGACAGCACCTTCTCTGGATTCATCCTCCTGGGCTTCTCCAACAGGCCTCAGCTGGAAAC
AGCTCTCTTTGTGGTCATCTTGATCATCTACTTTCTGAGCTTTCTGGGCAATGGCACCATTATACTTTTATCCATTGT
AGATCCTCGCCTCCATACCCCTATGTATTTCTTCCTCTCCAATCTCTCTTTTATGGATCTTTGTTTGACCACTTGTAC
TGTCCCTCAGACACTGGTCAACTTTAAGGGGAAGGACAAGACCATCACCTATGGTGGCTGCGTGACCCAGCTATTCAT
TGCCTTGGGACTCGG**GGGGGGGAG**TGGAGTGTGTCTTATTGTCTGCCATGGCCTATGACCGCTATGCAGCCGTCTGCCG
CCCACTCCACTACATGGTGAGCATGCATCCCCAACTTTGCTTGCAGTTGGTTGTAACCACTTGGCTCACAGGGTTTGG
CAATTCTGTGATACAGACAGCATTGACCATGACTCTCCCCCTCTGTGATAAAAACCAAGTGGATCATTTCTTCTGTGA
AGTTCCAGTGATGCTGAAACTGTCCTGCACCAACACCTCCATCAACGAGGCTGAAATCTTTGCTGTCAGTGTCTTCTT
CTTGGTGGTGCCTCTCTCACTCATCTTAGCATCCTATGGTCACATTACTCATGCAGTCCTGAAGATAAAGTCAGCTCA
AGGGAGGCAGAAGGCTTTTGGAACCTGTGGTTCTCACCTCCTGGTAGTGATCATTTTCTTTGGGACACTCATCTCCAT
GTACCTCCAGCCTCCCTCCAGTTATTCACAGGATGTGAACAAAAGCATTGCACTCTTCTATACTCTGGTGACTCCTCT
ACTGAATCCCCTAATTTACACTCTGAGGAACAAGGAAGTCAAAGGGGCAACTAAGAAGACTAGTGGGGAGGACCATAG
ATGCATGAGAAAGTTAACGCAGGGTTTGCAGTTCCAAACATTTGTGCACTAGAAGACTGC**TGA**GAAGCTACAAACTA

5 The disclosed nucleic acid sequence has 615 of 939 bases (65%) identical to a *Homo sapiens* olfactory receptor-like protein (OR2C1) gene (GENBANK-ID: AF098664) (E value = $1.7e\text{-}^{67}$).

The GPCR7 protein encoded by SEQ ID NO:28 has 327 amino acid residues, and is presented using the one-letter code in Table 7B (SEQ ID NO:29). The SignalP, Psort and/or
10 Hydropathy profile for GPCR7 predict that GPCR7 has a signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.6000. The SignalP shows a signal sequence is coded for in the first 44 amino acids, *i.e.*, with a cleavage site at the slash in the sequence GNG-TI, between amino acids 43 and 44. This is typical of this type of membrane protein.

15

Table 7B. Encoded GPCR7 protein sequence (SEQ ID NO:29).

MERANDSTFSGFILLGFSNRPQLETALFVVILIIYFLSFLGNG/TIILLSIVDPRLHTPMYFFLSNL
SFMDLCLTTCTVPQTLVNFKGKDKTITYGGCVTQLFIALGLGGGVECVLLSAMAYDRYAAVCRPLHY
MVSMHPQLCLQLVVTTWLTGFGNSVIQTALTMTLPLCDKNQVDHFFCEVPVMLKLSCTNTSINEAEI
FAVSVFFLVVPLSLILASYGHITHAVLKIKSAQGRQKAFGTCGSHLLVVIIFFGTLISMYLQPPSSY
SQDVNKSIALFYTLVTPLLNPLIYTLRNKEVKGATKKTSGEDHRCMRKLTQGLQFQTFVH

Table 7C. Patp alignments of GPCR7

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob. P(N)
patp:B43266 Human ORFX ORF3030 polypeptide sequence SE...	+2	927	2.6e-92
patp:B42796 Human ORFX ORF2560 polypeptide sequence SE..	+2	852	2.3e-84
patp:Y83390 Olfactory receptor protein OLF-5 - H. sapiens...	+2	727	4.1e-71
patp:Y54326 Amino acid sequence of marmot olfactory re..	+2	717	4.7e-70
patp:W75960 Human olfactory OLRCC15 receptor - H. sapiens...	+2	713	1.2e-69
patp:R27868 Odorant receptor clone F5 - Rattus rattus,..	+2	712	1.6e-69
patp:Y90874 Human G protein-coupled receptor GTAR14-5 ..	+2	697	6.2e-68

Further BLAST analysis produced the significant results listed in Table 7D. The disclosed GPCR7 protein (SEQ ID NO:29) has good identity with a number of olfactory receptor proteins.

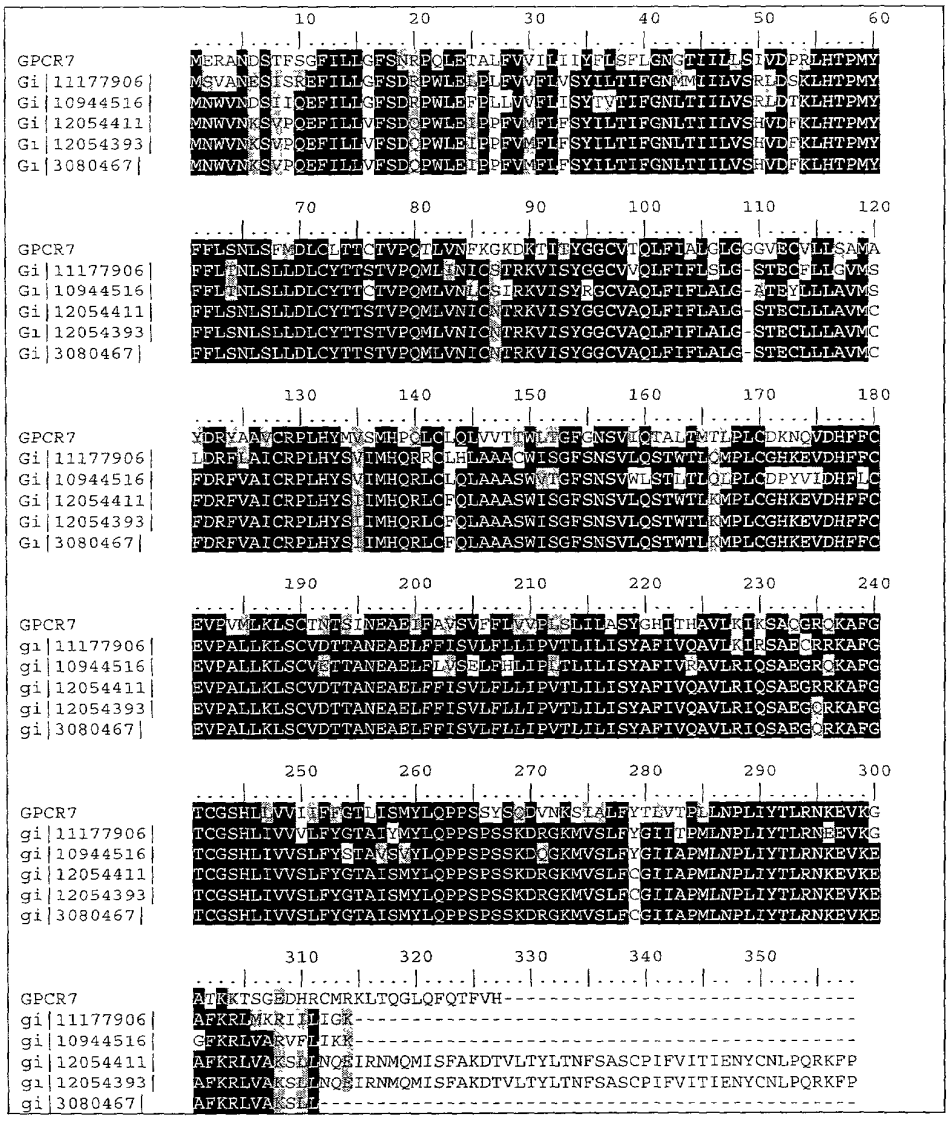
Table 7D. BLAST results for GPCR7

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|11177906|ref|NP_068632.1| (L34074)	OR *Rattus norvegicus*	313	168/304 (55%)	209/304 (68%)	2e-86
gi|10944516|emb|CAC14158.1| (AL133267) dJ408B20.2	novel 7 TM -OR family (hS6M1-32) *Homo sapiens*	313	169/304 (55%)	207/304 (67%)	2e-85
gi|12054411|emb|CAC20513.1| (AJ302593)	OR 41, K11 *Homo sapiens*	357	166/304 (54%)	206/304 (67%)	1e-83
gi|12054393|emb|CAC20504.1| (AJ302584 - 592)	OR *Homo sapiens*	357	165/304 (54%)	206/304 (67%)	4e-83
gi|3080467|emb|CAB11427.1| (Z98744)	OR *Homo sapiens*	310	165/304 (54%)	206/304 (67%)	4e-83

This information is presented graphically in the multiple sequence alignment given in Table 7E (with GPCR7 being shown on line 1) as a ClustalW analysis comparing GPCR7 with related protein sequences.

Table 7E. Information for the ClustalW proteins:

1) GPCR7 (SEQ ID NO:29)
2) gi|11177906|ref|NP_068632.1| Olfactory receptor Rattus norvegicus (SEQ ID NO:73)
3) gi|10944516|emb|CAC14158.1| dJ408B20.2 (novel 7 TM receptor (olfactory family) (hS6M1-32) Homo sapiens (SEQ ID NO:74)
4) gi|12054411|emb|CAC20513.1| olfactory receptor Homo sapiens (SEQ ID NO:75)
5) gi|12054393|emb|CAC20504.1| olfactory receptor Homo sapiens (SEQ ID NO:76)
6) gi|3080467|emb|CAB11427.1| olfactory receptor Homo sapiens (SEQ ID NO:77)



The presence of identifiable domains in GPCR7 was determined by searches using algorithms such as PROSITE, Blocks, Pfam, ProDomain, Prints and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (http:www.ebi.ac.uk/interpro/).

DOMAIN results for GPCR7 were collected from the Conserved Domain Database (CDD) with Reverse Position Specific BLAST. This BLAST samples domains found in the Smart and Pfam collections. The results are listed in Table 7F with the statistics and domain description. The results indicate that this protein contains the following protein domains (as defined by Interpro) at the indicated positions: domain name 7tm_1 (InterPro) 7 transmembrane receptor (rhodopsin family). Residues 61-142 of 7tm_1 (SEQ ID NO:45) are aligned with

GPCR7 41-180 (E = 5e-18) in Table 7F. Residues 307-377 of 7tm_1 also have identity with residues 222-291 (E=0.001) of GPCR7. This indicates that the sequence of GPCR7 has properties similar to those of other proteins known to contain this/these domain(s) and similar to the properties of these domains.

Table 7F. DOMAIN results for GPCR7

gnl|Pfam|pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family) Length = 377 (SEQ ID NO:45).

Score = 85.1 bits (209), Expect = 5e-18

```
                 ....|....|....|....|....|....|....|....|....|....|....|....|
GPCR7    GNGTTIILLSIVDPRLHTPMYFFLSNLSFMDICLTTCTVPQTLVNFKGKDKTITYGGCVTQ
TM7      GNVLVCMAVSREKALQTTTNYLIVSLAVADLLVATLVMPWVVYLEVVGEWKFSRIHCDIF

                 ....|....|....|....|....|....|....|....|....|....|....|....|
GPCR7    LFIALGLGGGVECVLLSAMAYDRYAAVCRPLHYMVSM-HPQLCLQLVVTTWLTGFGNSVI
TM7      VTLDVMMCTASILNLCAISIDRYTAVAMPMLYNTRYSSKRRVTVMIAIVWVLSFTISCPM

                   ..|....|....|....|....|....|....|....|....| ...|....|....|
GPCR7    QTALTMTLP--L--------CDKNQVDHFFC----------------------------
TM7      LFGLNNTDQNE--------CIIANP--------AFVVYSSIVSFYV----PFIVTLLVYI
```

The similarity information for the GPCR7 protein and nucleic acid disclosed herein suggest that GPCR7 may have important structural and/or physiological functions characteristic of the Olfactory Receptor family and the GPCR family. Therefore, the nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), and (v) a composition promoting tissue regeneration in vitro and in vivo (vi) biological defense weapon. The novel nucleic acid encoding GPCR7, and the GPCR7 protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

The nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in used in the treatment of infections such as bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to Neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease; multiple sclerosis; and treatment of Albright hereditary ostoeodystrophy, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy, and psychotic and neurological disorders, including anxiety,

schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome and/or other pathologies and disorders.

The disclosed GPCR7 polypeptides can be used as immunogens to produce antibodies specific for the invention, and as vaccines. They can also be used to screen for potential agonist and antagonist compounds. For example, a cDNA encoding the GPCR-like protein may be useful in gene therapy, and the GPCR-like protein may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease; multiple sclerosis; and treatment of Albright hereditary ostoeodystrophy, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy, and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome and/or other pathologies and disorders. The novel nucleic acid encoding GPCR-like protein, and the GPCR-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. For example the disclosed GPCR7 protein has multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated GPCR7 epitope is from about amino acids 5 to 15. In another embodiment, a GPCR7 epitope is from about amino acids 50 to 65. In additional embodiments, GPCR7 epitopes are from amino acids 75 to 100, 125-140, 160-170, 220 to 245, 255-275 and from amino acids 290 to 330. These novel proteins can also be used to develop assay system for functional analysis. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-GPCRX Antibodies" section below.

GPCR8

A novel nucleic acid was identified on chromosome 6 by TblastN using CuraGen Corporation's sequence file for GPCR probes or homologs and run against the Genomic Daily

Files made available by GenBank. The nucleic acid was further predicted by the program GenScan™, including selection of exons. These were further modified by means of similarities using BLAST searches. The sequences were then manually corrected for apparent inconsistencies, thereby obtaining the sequences encoding the full-length protein. The disclosed novel GPCR8 nucleic acid of 958 nucleotides (also referred to as dj408b20_A) is shown in Table 8A. An open reading begins with an ATG initiation codon at nucleotides 4-6 and ends with a TGA codon at nucleotides 955-957. A putative untranslated region upstream from the initiation codon and downstream from the termination codon are underlined in Table 8A, and the start and stop codons are in bold letters.

Table 8A. GPCR8 Nucleotide Sequence (SEQ ID NO:30)

GCA**ATG**GAAAAATCCAATGTCAGCTCAGTGTATGGTTTTATCTTGGTGGGTTTCTCTGATCGTCCCAAGCTG
GAGATGGTGCTCTTTACAGTAAATTTTATTCTGTATTCAGTGGCTGTGCTGGGAAATTCAACCATAATCC
TTGTGTGTATATTAGACTCTCAACTTCATACCCCAATGTACTTCTTTCTGGCAAATCTTTCCTTTCTAGA
TCTCTGCTTCAGTACTAGTTGCATCCCACAAATGCTGGTAAACCTCTGGGGCCCTGACAAGACTATTAGC
TGTGCTGGCTGTGTTGTCCAGCTTTTCTCTTTCCTTTCTGTCAGGGGAATTGAGTGCATCCTTCTGGCTG
TCATGGCCTATGACAGCTATGCTGCAGTCTGCAAACCGTTGCGCTATCTGGTCATTATGCACCTCCAGCT
GTGTCTAGGACTGATGGCTGCAGCCTGGGGGAGTGGACTGGTCAATGCCGTTGTCATGTCACCACTAACA
ATGACCCTCTCCAGAAGTGGCCGCCGCCGAGTTAACCATTTCCTCTGTGAAAAGCCAGCACTGATCAAGA
TGGCTTGTTTGGATGTTCGTGCAGTGGAAATGCTGGCTTTTGCTTTTGCCGTTCTCATTGTCCTACTGCC
CCTCACTCTTATTCTTGTCTCCTACGGCTACATTGCTGCAGCTGTGCTAAGCATCAAGTCAGCTGCCAGG
CAATGGAAGGCCTTCCATACCTGTAGCTCTCACCTCACAGTGGTCTCCCTGTTTTATGGGAGCATCATCT
ATATGTATATGCAGCCAGGAAACAGTTCTTCCCAAGACCAAGGCAAGTTTCTCACTCTCTTCTACAACCT
GGTGACTCCTATGTTGAATCTGCTCATCTATACTTTAAGGAATAAGGAGGTGAAAGGAGCACTGAAGAAG
GTTTTGGGGAGGCAAAATGAACTGGAGAAATATGATAAGTTG**TGA**A

The disclosed nucleic acid sequence has 768 of 1148 bases (66%) identical to a *Homo sapiens* OR-like (gb:GENBANK-ID:HS88J8|acc:AL035402 Human DNA sequence from clone 88J8 on chromosome 6p21.31-21.33, which contains a gene for a novel 7 transmembrane receptor (rhodopsin family) (olfactory receptor like) protein, a pseudogene similar to olfactory receptor genes and a GTP binding protein SARA (mouse) pseudogene. The clone also contains ESTs, an STS and GSS, complete sequence from *Homo sapiens*, 47216 bp.) (E value = $1.3e\text{-}^{81}$).

The GPCR8 protein encoded by SEQ ID NO:30 has 317 amino acid residues, and is presented using the one-letter code in Table 8B (SEQ ID NO:31). The SignalP, Psort and/or Hydropathy profile for GPCR8 predict that GPCR8 has a signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.6850. The SignalP shows a signal sequence is coded for in the first 42 amino acids, *i.e.*, with a cleavage site at the slash in the sequence VLG-NS, between amino acids 41 and 42. This is typical of this type of membrane protein.

Table 8B. Encoded GPCR8 protein sequence (SEQ ID NO:31).
MEKSNVSSVYGFILVGFSDRPKLEMVLFTVNFILYSVAVLG/NSTIILVCILDSQLHTPMYFFLANLSF LDLCFSTSCIPQMLVNLWGPDKTISCAGCVVQLFSFLSVRGIECILLAVMAYDSYAAVCKPLRYLVIMH LQLCLGLMAAAWGSGLVNAVVMSPLTMTLSRSGRRRVNHFLCEKPALIKMACLDVRAVEMLAFAFAVLI VLLPLTLILVSYGYIAAAVLSIKSAARQWKAFHTCSSHLTVVSLFYGSIIYMYMQPGNSSSQDQGKFLT LFYNLVTPMLNLLIYTLRNKEVKGALKKVLGRQNELEKYDKL

The full amino acid sequence of the protein of the invention was found to have 187 of 305 amino acid residues (61%) identical to, and 239 of 305 residues (78%) positive with, the 320 amino acid residue novel transmembrane receptor (rhodopsin family, OR-like, HS6M1-15) protein from *Homo sapiens* (ptnr:SPTREMBL-ACC:Q9Y3N9) (E value = 2.4e-^{101}).

Patp results include those listed in Table 8C.

Table 8C. Patp alignments of GPCR8

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob. P(N)
patp:B43266 Human ORFX ORF3030 polypeptide sequence SE. ..	+1	878	4.1e-87
patp:Y54326 Amino acid sequence of marmot olfactory receptor.	+1	814	2.5e-80
patp:B42796 Human ORFX ORF2560 polypeptide sequence SE.. .	+1	807	1.4e-79
patp:Y83390 Olfactory receptor protein OLF-5 - H, sapiens.. .	+1	652	3.6e-63

For example, a BLAST against Y54326, a 237 amino acid olfactory receptor protein from *Marmota marmota*, produced 156/237 (65%) identity, and 188/237 (79%) positives (E = 2.5e-80), with long segments of amino acid identity, as shown in Table 8D. WO 99/67282.

Table 8D. Alignment of GPCR8 with Y54326 (SEQ ID NO:88)

```
                10        20        30        40        50        60
        ....|....|....|....|....|....|....|....|....|....|....|....|
Y54326  ------------------------------------------------------PMY 3
GPCR8   MEKSNVSSVYGFILVGFSDRPKLEMVLFTVNFILYSVAVLGNSTIILVCILDSQLHTPMY 60

                70        80        90        100       110       120
        ....|....|....|....|....|....|....|....|....|....|....|....|
Y54326  LFLGNLSFLDLSFTSSIPQLLHNLSGRDKTISYVGCVVQLFLFLGLGGVECLLLAVMAY 63
GPCR8   FFLANLSFLDLCFSTSCIPQMLVNLWGPDKTISCAGCVVQLFSFLSVRGIECILLAVMAY 120

                130       140       150       160       170       180
        ....|....|....|....|....|....|....|....|....|....|....|....|
Y54326  DRFVAVCKPLHYTVIMSSRLCLGLVSVAWGCGMANSLVMSPVTLQLPFCGHNKVDHFLCE 123
GPCR8   DSYAAVCKPLRYLVIMHLQLCLGLMAAAWGSGLVNAVVMSPLTMTLSRSGRRRVNHFLCE 180

                190       200       210       220       230       240
        ....|....|....|....|....|....|....|....|....|....|....|....|
Y54326  MPALIRMACVNTVAIEGTVFVLAVGIVLSPLVFILVSYGHIVRAVFRIQSSSGRHRIFNT 183
GPCR8   KPALIKMACLDVRAVEMLAFAFAVLIVLLPLTLILVSYGYIAAAVLSIKSAARQWKAFHT 240

                250       260       270       280       290       300
        ....|....|....|....|....|....|....|....|....|....|....|....|
Y54326  CGSHLTVVSLFYGNIIYMYMQPGSRSSQDQGKFLTLFYNIVTPILNPFIYSLRN------ 237
GPCR8   CSSHLTVVSLFYGSIIYMYMQPGNSSSQDQGKFLTLFYNLVTPMLNLLIYTLRNKEVKGA 300
```

Further BLAST analysis produced the significant results listed in Table 8E. The disclosed GPCR8 protein (SEQ ID NO:31) has good identity with a number of olfactory receptor proteins.

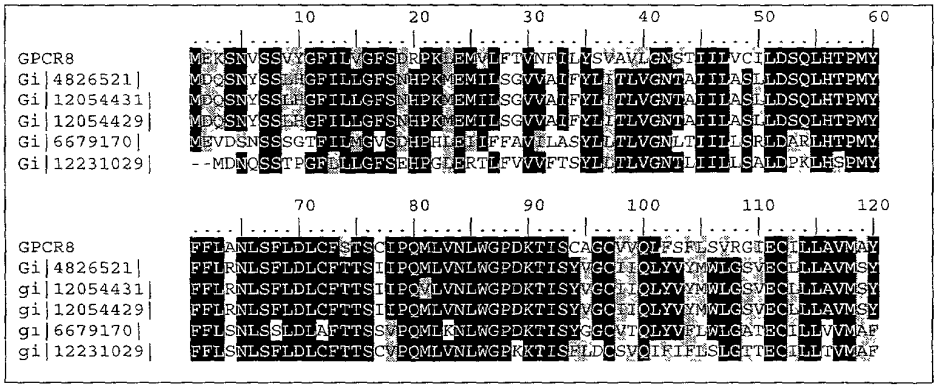
Table 8E. BLAST results for GPCR8

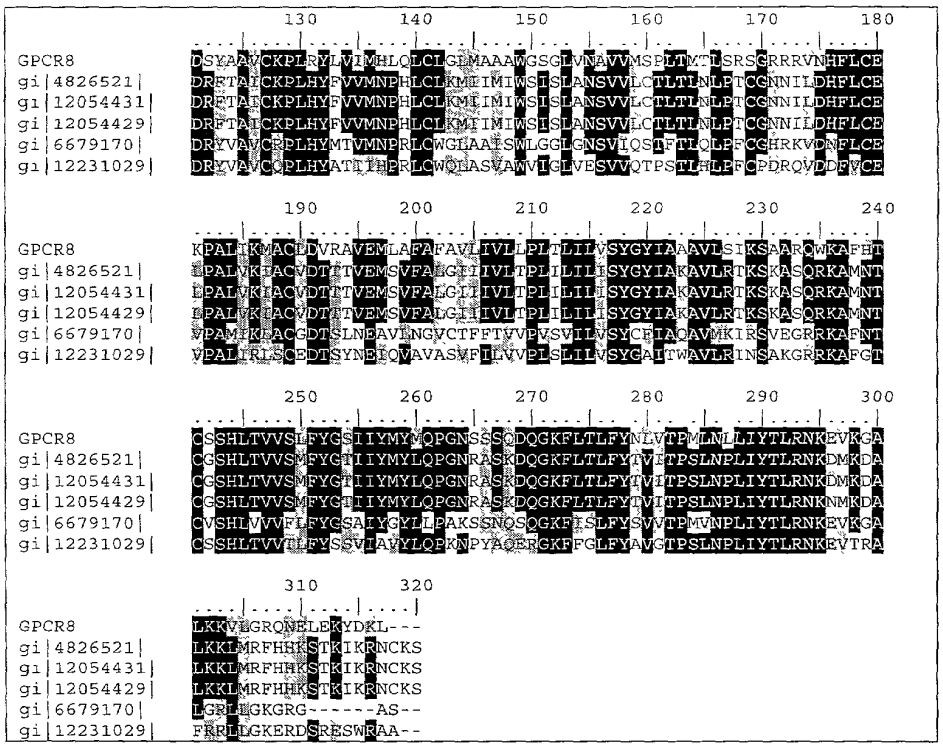
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
Gi|4826521|emb|CAB42853.1| (AL035402, dJ88J8.1) (AJ302594-99) (AJ302600-01)	novel 7 tm receptor protein (rhodopsin fam., OR-like) (hs6M1-15) *Homo sapiens*	320	177/305 (58%)	226/305 (74%)	5e-92
Gi|12054431|emb|CAC20523.1| (AJ302603)	OR *Homo sapiens*	320	176/305 (57%)	226/305 (73%)	1e-91
Gi|12054429|emb|CAC20522.1| (AJ302602)	OR *Homo sapiens*	320	177/305 (58%)	225/305 (73%)	1e-91
Gi|6679170|ref|NP_032788.1|	OR 15 (OR3) *Mus musculus*	312	166/307 (54%)	211/307 (68%)	2e-81
Gi|12231029|sp|Q15062|O2H3_HUMAN	OR 2H3 *Homo sapiens*	316	163/306 (53%)	208/306 (67%)	5e-81

This information is presented graphically in the multiple sequence alignment given in Table 8F (with GPCR8 being shown on line 1) as a ClustalW analysis comparing GPCR8 with related protein sequences.

Table 8F. Information for the ClustalW proteins:

1) GPCR8 (SEQ ID NO:31)
2) gi|4826521|emb|CAB42853.1| dJ88J8.1 (novel 7 TM receptor (rhodopsin family) (OR like) protein) (hs6M1-15)) Homo sapiens (SEQ ID NO:59)
3) gi|12054431|emb|CAC20523.1| olfactory receptor Homo sapiens (SEQ ID NO:60)
4) gi|12054429|emb|CAC20522.1| olfactory receptor Homo sapiens (SEQ ID NO:61)
5) gi|6679170|ref|NP_032788.1| olfactory receptor 15 Mus musculus (SEQ ID NO:58)
6) gi|12231029|sp|Q15062|O2H3_HUMAN OR 2H3 (OR-LIKE PRT FAT11) (SEQ ID NO:56)



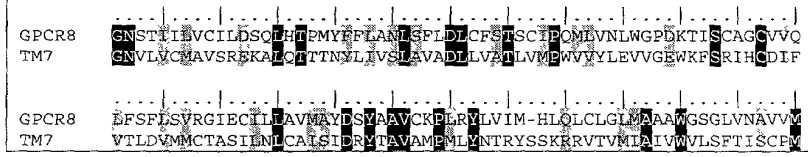


The presence of identifiable domains in GPCR8 was determined by searches using algorithms such as PROSITE, Blocks, Pfam, ProDomain, Prints and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (http:www.ebi.ac.uk/interpro/).

DOMAIN results for GPCR8 were collected from the Conserved Domain Database (CDD) with Reverse Position Specific BLAST. This BLAST samples domains found in the Smart and Pfam collections. The results are listed in Table 8G with the statistics and domain description. The results indicate that GPCR8 contains the 7tm_1 (InterPro) 7 transmembrane receptor (rhodopsin family)domain (as defined by Interpro) at amino acid positions residues 41-170. This indicates that the sequence of GPCR8 has properties similar to those of other proteins known to contain this/these domain(s) and similar to the properties of these domains.

Table 8G. DOMAIN results for GPCR8

gnl|Pfam|pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family)
Length = 377 (SEQ ID NO:45).Score = 85.9 bits (211), Expect = 3e-18



The similarity information for the GPCR8 protein and nucleic acid disclosed herein suggest that GPCR8 may have important structural and/or physiological functions characteristic of the Olfactory Receptor family and the GPCR family. Therefore, the nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), and (v) a composition promoting tissue regeneration in vitro and in vivo (vi) biological defense weapon. The novel nucleic acid encoding GPCR8, and the GPCR8 protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

The disclosed GPCR8 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer; immune response; AIDS; asthma; Crohn's disease; multiple sclerosis; and treatment of Albright hereditary ostoeodystrophy and/or other pathologies and disorders. For example, a cDNA encoding the GPCR-like protein may be useful in gene therapy, and the GPCR-like protein may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer; immune response; AIDS; asthma; Crohn's disease; multiple sclerosis; and treatment of Albright hereditary ostoeodystrophy. The novel nucleic acid encoding GPCR-like protein, and the GPCR-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These materials are further useful in the generation of antibodies that bind immuno-specifically to the novel GPCR8 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-GPCRX Antibodies" section below. For example the disclosed GPCR8 protein has multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated GPCR8 epitope is from about amino acids 160 to 180. In another embodiment, a GPCR8 epitope is from about amino acids 225 to 240. In additional embodiments, GPCR8 epitopes are from amino acids 260 to 275 and from amino acids 290 to 330. These novel proteins can also be used to develop assay system for functional analysis.

GPCR9

A novel nucleic acid was identified on chromosome 11 by TblastN using CuraGen Corporation's sequence file for GPCR probe or homolog, run against the Genomic Daily Files made available by GenBank. The nucleic acid was further predicted by the program GenScan™, including selection of exons. These were further modified by means of similarities using BLAST searches. The sequences were then manually corrected for apparent inconsistencies, thereby obtaining the sequences encoding the full-length protein. The disclosed novel GPCR9 nucleic acid of 946 nucleotides (also referred to as 6-L-19-B) is shown in Table 9A. An open reading begins with an ATG initiation codon at nucleotides 5-7 and ends with a TAA codon at nucleotides 932-934. A putative untranslated region upstream from the initiation codon and downstream from the termination codon are underlined in Table 9A, and the start and stop codons are in bold letters.

In a search of sequence databases, it was found, for example, that the nucleic acid sequence has 563 of 902 bases (62%) identical to (E= 8.5e-47) a *Mus musculus* gene for odorant receptor A16 (GENBANK-ID: GENBANK-ID:AB030896|acc:AB030896). In a search of sequence databases, partial matches (353 of 373 bases, 94% identical) were also identified with the nucleotides 1- 372 of GPCR9 identical with nucleotides 306-678 of a *Homo sapiens* GPCR EST (GENBANK-ID: GENBANK-ID: AW182678|acc:AW182678; xj45d11.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2660181 3' similar to TR:Q9Z1V0 Q9Z1V0 OLFACTORY RECEPTOR C6). This 94% match (E=1.1e-69) between regions of the public sequence and regions of the present invention (gene) suggests that the present invention (gene) could be a splice variant of the public GPCR EST (partial mRNA). This also supports identification of GPCR9 as a GPCR. In a search of sequence databases, partial matches (94 of 100 bases, 94% identical) of nucleotides 893-794 of GPCR9 with nucleotides 251-348 of a *Homo sapiens* GPCR EST (GENBANK-ID: AA206680|acc:AA206680: zq51c11.r1 Stratagene neuroepithelium (#937231) Homo sapiens cDNA clone IMAGE:645140 5' similar to contains L1.b2 L1 repetitive element). This 94% match between nucleotides of the public sequence and nucleotides of the GPCR9 sequence suggests that GPCR9 may be a splice variant of the public GPCR EST (partial mRNA).

Table 9A. GPCR9 Nucleotide Sequence (SEQ ID NO:32)

GTTC**ATG**GAAAATAGGAATATTGTCACTGTCTTTATTCTCCTGGGACTTTCTCAAAACAAGAACATTGAA
GTTTTTTGGTTTGTATTATTTGTATTTTGCTACATTGCTATTTGGATGGAAAACTTCATCATAATGATTT
CTATCATGTACATTTGGCTAATTGACCAACCCATGTATTTCTTCCTTAATTACCTCGCACTCTCAGATCT
TTGCTACATATCCACTGTGGCCCCCAAGCTAATGATTGACCTACTAACAGAAAGGAAGATCGTTTCCTAT
AATAACTGCATGATACAGCTATTTATCACTCACTTCCTTGGAGACATTGAGATCTTCATACTCAAAGCAA
TGGCCTATGACCACTACATAGCCATCTGCAAGCACCTGCACTACACCATCATCACGACCAAGCAAAGCTG
TAACACCATCATCATAGCTTGTTGTACTGGGGGATTTATACACTCTGCCAGTCAGTTTCTTCTTACCATC
TTCTTACCGTTCTGTGGTCTTAATGAGATAGATCAGTACTTCTGCTATGTGTATCCTCTGCTGAAGTTGG
CTCGCATTGATATATACAGAATTGGTTTCTTGGTAATTGTTAATTCAGGCCTGATTTCTTTGTTGGCTTT
TGTGATTTTGATGGTGTCTTATTATTTGATATTATCCACCATCAGGGTTTACTCTGCTGAGAGTCATACC
AAAGCTCTTTCAACCTGTAGCTCTCACATAATAGTTGTGGTCCTATTCTTTGTGCCTGCCCTCTTCATTT
ACATCAGACCAGCCATAACTTTTCCAGAAGATAAAGTGTTTGTTCTCTTCTGTGCCATCATTGCTCCCAT
GTTCAGTCTTCTTATCTACATGCTGAGAAAGGTGGAGATGAAGAACGCTGTAAGGAAAATGTGGTGTCAT
CAATTGCTTCTGGCAAGGAAG**TAA**CTTGTATGAAAG

The GPCR9 protein encoded by SEQ ID NO:32 has 309 amino acid residues, and is presented using the one-letter code in Table 9B (SEQ ID NO:33). The SignalP, Psort and/or Hydropathy profile for GPCR9 predict that GPCR9 has a signal peptide and is likely to be localized at the endoplasmic reticulum membrane with a certainty of 0.6850 or to the plasma membrane with a certainty of 0.6400. The SignalP predicts a cleavage site at the sequence IWM-EN between amino acids 38 and 39 as indicated by the slash in Table 9B.

Table 9B. Encoded GPCR9 protein sequence (SEQ ID NO:33)

MENRNIVTVFILLGLSQNKNIEVFWFVLFVFCYIAIWM/ENFIIMISIMYIWLIDQPMYFFLNYLALSDLC
YISTVAPKLMIDLLTERKIVSYNNCMIQLFITHFLGDIEIFILKAMAYDHYIAICKHLHYTIITTKQSCN
TIIIACCTGGFIHSASQFLLTIFLPFCGLNEIDQYFCYVYPLLKLARIDIYRIGFLVIVNSGLISLLAFV
ILMVSYYLILSTIRVYSAESHTKALSTCSSHIIVVVLFFVPALFIYIRPAITFPEDKVFVLFCAIIAPMF
SLLIYMLRKVEMKNAVRKMWCHQLLLARK

The full amino acid sequence of the protein of the invention was found to have 140 of 300 amino acid residues (46%) identical to, and 193 of 300 residues (64%) positive with, the 302 amino acid residue odorant receptor A16 protein from *Mus musculus* (ptnr:TREMBLNEW-ACC:BAA86127) (E value = $1.0e^{-72}$).

Patp results include those listed in Table 9C.

Table 9C. Patp alignments of GPCR9

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob. P(N)
patp:Y83390 Olfactory receptor protein OLF-5 - H. sapiens...	+1	801	5.9e-79
patp:Y83394 Olfactory receptor protein OLF-9 - H. sapiens...	+1	791	6.8e-78
patp:Y90877 Human G protein-coupled receptor GTAR11-3 ...	+1	766	3.0e-75
patp:Y90877 Human G protein-coupled receptor GTAR11-3 ...	+1	766	3.0e-75
patp:Y90875 Human G protein-coupled receptor GTAR11-1 ...	+1	753	7.2e-74
patp:Y90875 Human G protein-coupled receptor GTAR11-1 ...	+1	753	7.2e-74
patp:Y83387 Olfactory receptor protein OLF-2 - H. sapiens...	+1	741	1.3e-72

For example, a BLAST against Y83390, a 305 amino acid olfactory receptor protein from *Homo sapiens*, produced 158/305 (51%) identity, and 213/305 (69%) positives (E = 5.9e-79), with long segments of amino acid identity. WO 00/21985.

Further BLAST analysis produced the significant results listed in Table 9D. The disclosed GPCR9 protein (SEQ ID NO:33) has good identity with a number of olfactory receptor proteins.

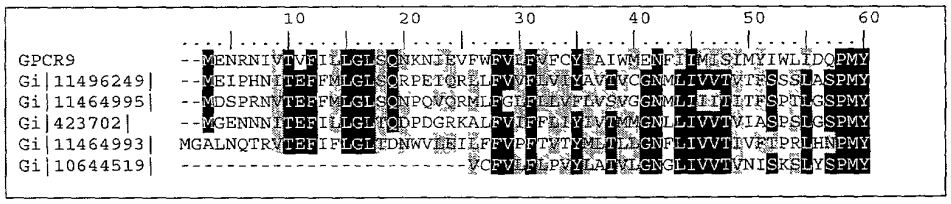
Table 9D. BLAST results for GPCR9

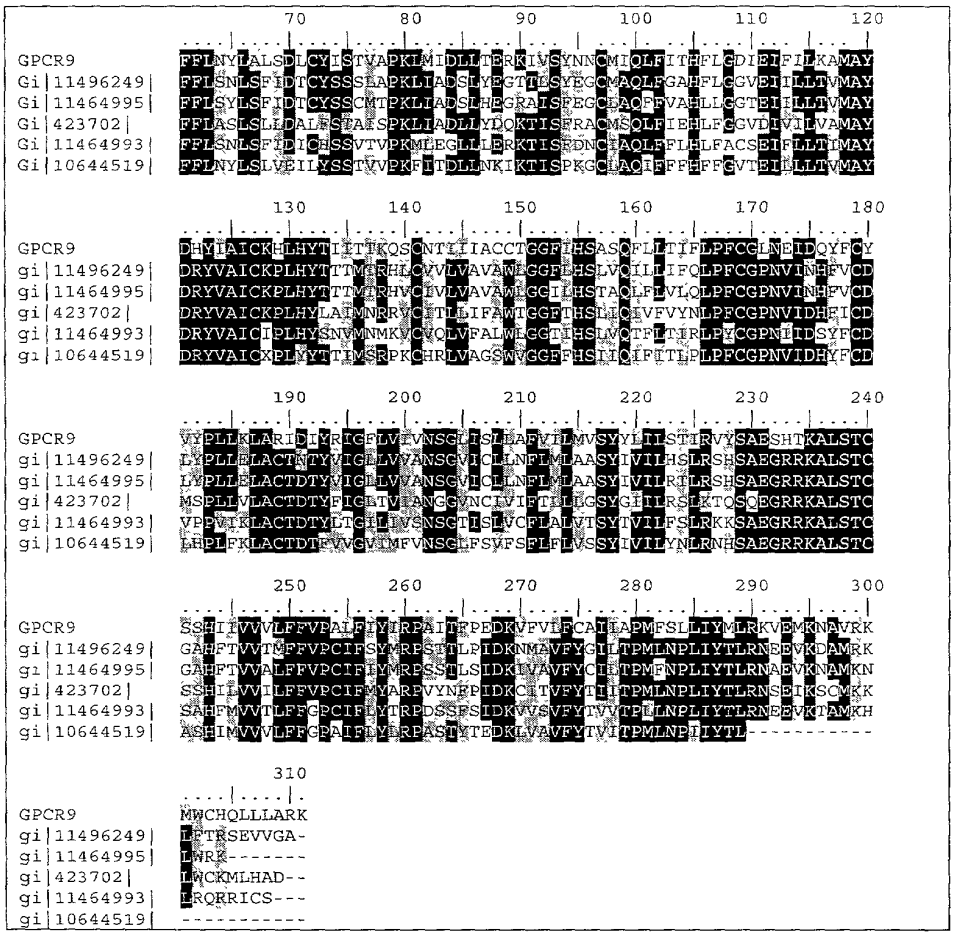
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
Gi|11496249|ref|NP_ 067343.1| (AB030895)	Odorant receptor MOR18 *Mus musculus*	308	140/300 (46%)	185/300 (61%)	1e-59
Gi|11464995|ref|NP_ 065261.1| AB030896)	Odorant receptor A16 *Mus musculus*	302	137/300 (45%	185/300 (61%)	2e-59
Gi|423702|pir||S297 10	Olfactory receptor OR18 - rat	307	142/303 (46%)	186/303 (60%)	6e-57
Gi|11464993|ref|NP_ 065260.1| (AB030894)	Odorant receptor MOR83 *Mus musculus*	308	133/297 (44%)	182/297 (60%)	7e-55
Gi|10644519|gb|AAG2 1324.1|AF271051_1 (AF271051)	Odorant receptor Mus musculus	264	124/262 (47%)	169/262 (64%)	6e-51

This information is presented graphically in the multiple sequence alignment given in Table 9E (with GPCR9 being shown on line 1) as a ClustalW analysis comparing GPCR9 with related protein sequences.

Table 9E. Information for the ClustalW proteins:

1) GPCR9 (SEQ ID NO:33)
2) gi|11496249|ref|NP_067343.1| odorant receptor 16 Mus musculus (SEQ ID NO:78)
3) gi|11464995|ref|NP_065261.1| gene for odorant receptor A16 Mus musculus (SEQ ID NO:79)
4) gi|423702|pir||S29710 olfactory receptor OR18 - rat (SEQ ID NO:80)
5) gi|11464993|ref|NP_065260.1| gene for odorant receptor MOR83 Mus musculus (SEQ ID NO:81)
6) gi|10644519|gb|AAG21324.1|AF271051_1 odorant receptor Mus musculus (SEQ ID NO:70)





The presence of identifiable domains in GPCR9 was determined by searches using algorithms such as PROSITE, Blocks, Pfam, ProDomain, Prints and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (http:www.ebi.ac.uk/interpro/).

DOMAIN results for GPCR9 were collected from the Conserved Domain Database (CDD) with Reverse Position Specific BLAST. This BLAST samples domains found in the Smart and Pfam collections. The results are listed in Table 9F with the statistics and domain description. The results indicate that GPCR9 contains the 7tm_1 (InterPro) 7 transmembrane receptor (rhodopsin family)domain (as defined by Interpro) at amino acid positions residues 56-234. This indicates that the sequence of GPCR9 has properties similar to those of other proteins known to contain this/these domain(s) and similar to the properties of these domains.

Table 9F. DOMAIN results for GPCR9

gnl|Pfam|pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family) Length = 377 (SEQ ID NO:45).

Score = 73.6 bits (179), Expect = 1e-14

```
GPCR9          ----------PMYFFLNYLALSDLCYISTVAPKMIDLFTERKIVSYNNCMTQ
TM7    GNVLVCMAVSREKALQTTTNYLIVSLAVADLLVATLVMPWVVYLEVVGEWKFSRIHCDIF

GPCR9  LFITHFLGDIEIFILKAMAYDRYIAICKHLHY-TIITTKQSCNTIIIACCTGGFIHSASQ
TM7    VTFDVVMCTASILNLCAISIDRYTAVAMPMLYNTRYSSKRRVTVMIAIVWVLSFTISCPM

GPCR9  FLL-TIFLPFCGLNEIDQYFCYVYPLLKLARIDIYRIGFL---VIVNSGLISFLAFVILM
TM7    LFGLNNTDQNE--------CIIANP--------AFVVYSSIVSFYV----PFIVTLLVYI

GPCR9  VSYYLILSTIRVYSAESHTKA-------------------------------------
TM7    KIYIVLRRRRKRVNTKRSSRAFRANLKAPLKGNCTHPEDMKLCTVIMKSNGSFPVNRRRV
```

The similarity information for the GPCR9 protein and nucleic acid disclosed herein suggest that GPCR9 may have important structural and/or physiological functions characteristic of the Olfactory Receptor family and the GPCR family. Therefore, the nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), and (v) a composition promoting tissue regeneration in vitro and in vivo (vi) biological defense weapon. The novel nucleic acid encoding GPCR9, and the GPCR9 protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

The nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in used in the treatment of infections such as bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease; multiple sclerosis; and treatment of Albright hereditary ostoeodystrophy, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy, and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias,

such as Huntington's disease or Gilles de la Tourette syndrome and/or other pathologies and disorders.

The polypeptides can be used as immunogens to produce antibodies specific for the invention, and as vaccines. They can also be used to screen for potential agonist and antagonist compounds. For example, a cDNA encoding the GPCR-like protein may be useful in gene therapy, and the GPCR-like protein may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease; multiple sclerosis; and treatment of Albright hereditary ostoeodystrophy, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy, and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome and/or other pathologies and disorders. The novel nucleic acid encoding GPCR-like protein, and the GPCR-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

These materials are further useful in the generation of antibodies that bind immuno-specifically to the novel GPCR9 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-GPCRX Antibodies" section below. For example the disclosed GPCR9 protein has multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated GPCR9 epitope is from about amino acids 75 to 100. In another embodiment, a GPCR9 epitope is from about amino acids 115 to 145. In additional embodiments, GPCR9 epitopes are from amino acids 225 to 240 and from amino acids 290 to 310. These novel proteins can also be used to develop assay system for functional analysis.

GPCR10

GPCR10 includes a family of three similar nucleic acids and three similar proteins disclosed below. The disclosed nucleic acids encode GPCR, OR-like proteins.

GPCR10a

The disclosed novel nucleic acid was identified on chromosome 11 by TblastN using CuraGen Corporation's sequence file for GPCR probe or homolog, run against the Genomic Daily Files made available by GenBank. The nucleic acid was further predicted by the program GenScan™, including selection of exons. These were further modified by means of similarities using BLAST searches. The sequences were then manually corrected for apparent inconsistencies, thereby obtaining the sequences encoding the full-length protein. GPCR10a is a 948 bp long nucleic acid (also referred to as 6-L-19-A) as shown in Table 10A (SEQ ID NO:34). An ORF begins with an ATG initiation codon at nucleotides 7-9 and ends with a TAA codon at nucleotides 934-936. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in Table 10A, and the start and stop codons are in bold letters.

Table 10A. GPCR10a Nucleotide Sequence (SEQ ID NO:34)

TGAGAA**ATG**GAAAATCAAAACAATGTGACTGAATTCATTCTTCTGGGTCTCACAGAGAACCTGGAGCTGT
GGAAAATATTTTCTGCTGTGTTTCTTGTCATGTATGTAGCCACAGTGCTGGAAAATCTACTTATTGTGGT
AACTATTATCACAAGTCAGAGTCTGAGGTCACCTATGTATTTTTTTCTTACCTTCTTGTCCCTTTTGGAT
GTCATGTTCTCATCTGTCGTTGCCCCCAAGGTGATTGTAGACACCCTCTCCAAGAGCACTACCATCTCTC
TCAAAGGCTGCCTCACCCAGCTGTTTGTGGAGCATTTCTTTGGTGGTGTGGGGATCATCCTCCTCACTGT
GATGGCCTATGACCGCTACGTGGCCATCTGTAAGCCCCTGCACTACACGATCATCATGAGTCCACGGGTG
TGCTGCCTAATGGTAGGAGGGGCTTGGGTGGGGGGATTTATGCACGCAATGATACAACTTCTCTTCATGT
ATCAAATACCCTTCTGTGGTCCTAATATCATAGATCACTTTATATGTGATTTGTTTCAGTTGTTGACACT
TGCCTGCACGGACACCCACATCCTGGGCCTCTTAGTTACCCTCAACAGTGGGATGATGTGTGTGGCCATC
TTTCTTATCTTAATTGCGTCCTACACGGTCATCCTATGCTCCCTGAAGTCTTACAGCTCTAAAGGGCGGC
ACAAAGCCCTCTCTACCTGCAGCTCCCACCTCACGGTGGTTGTATTGTTCTTTGTCCCCTGTATTTTCTT
GTACATGAGGCCTGTGGTCACTCACCCCATAGACAAGGCAATGGCTGTGTCAGACTCAATCATCACACCC
ATGTTAAATCCCTTGATCTATACACTGAGGAATGCAGAGGTGAAAAGTGCCATGAAGAAACTCTGGATGA
AATGGGAGGCTTTGGCTGGGAAA**TAA**CTGCAATGCTGA

The GPCR10a protein encoded by SEQ ID NO:34 has 309 amino acid residues, and is presented using the one-letter code in Table 10B (SEQ ID NO:35). The SignalP, Psort and/or Hydropathy profile for GPCR10a predict that GPCR10a has a signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.6000. The SignalP predicts a cleavage site at the sequence VLE-NL, between amino acids 39 and 40, as indicated by the slash in Table 10B.

In a search of sequence databases, it was found, for example, that the nucleic acid sequence has 625 of 908 bases (68%) identical to a 909 bp *Mus musculus* gene for odorant receptor A16 mRNA (GENBANK-ID: AB030896|acc:AB030896) (E= 3.0e-75).

Table 10B. Encoded GPCR10a protein sequence (SEQ ID NO:35).
MENQNNVTEFILLGLTENLELWKIFSAVFLVMYVATVLE/NLLIVVTIITSQSLRSPMYFFLTFLSLLDVM FSSVVAPKVIVDTLSKSTTISLKGCLTQLFVEHFFGGVGIILLTVMAYDRYVAICKPLHYTIIMSPRVCC LMVGGAWVGGFMHAMIQLLFMYQIPFCGPNIIDHFICDLFQLLTLACTDTHILGLLVTLNSGMMCVAIFL ILIASYTVILCSLKSYSSKGRHKALSTCSSHLTVVVLFFVPCIFLYMRPVVTHPIDKAMAVSDSIITPML NPLIYTLRNAEVKSAMKKLWMKWEALAGK

The full amino acid sequence of the protein of the invention was found to have 183 of 302 amino acid residues (60%) identical to, and 232of 302 residues (76%) positive with, the 307 amino acid residue OR18 odorant receptor protein from *Rattus sp.*(ptnr: TREMBLNEW-ACC:G264618).

GPCR10b

GPCR10a (6-L-19-A) was subjected to an exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such suitable sequences were then employed as the forward and reverse primers in a PCR amplification based on a wide range of cDNA libraries. The resulting amplicon was gel purified, cloned and sequenced to high redundancy to provide GPCR10b, which is also referred to as 6-L-19-A1.

The nucleotide sequence for GPCR10b (948 bp, SEQ ID NO:36) is presented in Table 10C. The nucleotide sequence differs from GPCR10a by one nucleotide change (numbered with respect to GPCR10a) T404 >C.

Table 10C. GPCR10b Nucleotide Sequence (SEQ ID NO:36)
TGAGAAATGGAAAATCAAAACAATGTGACTGAATTCATTCTTCTGGGTCTCACAGAGAACCTGGAGCTGTGGAAAATATT TTCTGCTGTGTTTCTTGTCATGTATGTAGCCACAGTGCTGGAAAATCTACTTATTGTGGTAACTATTATCACAAGTCAGA GTCTGAGGTCACCTATGTATTTTTTTCTTACCTTCTTGTCCCTTTGGATGTCATGTTCTCATCTGTCGTTGCCCCCAAG GTGATTGTAGACACCCTCTCCAAGAGCACTACCATCTCTCTCAAAGGCTGCCTCACCCAGCTGTTTGTGGAGCATTTCTT TGGTGGTGTGGGGATCATCCTCCTCACTGTGATGGCCTATGACCGCTACGTGGCCATCTGTAAGCCCCTGCACTACACGA TCACCATGAGTCCACGGGTGTGCTGCCTAATGGTAGGAGGGGCTTGGGTGGGGGGATTTATGCACGCAATGATACAACTT CTCTTCATGTATCAAATACCCTTCTGTGGTCCTAATATCATAGATCACTTTATATGTGATTTGTTTCAGTTGTTGACACT TGCCTGCACGGACACCCACATCCTGGGCCTCTTAGTTACCCTCAACAGTGGGATGATGTGTGTGGCCATCTTTCTTATCT TAATTGCGTCCTACACGGTCATCCTATGCTCCCTGAAGTCTTACAGCTCTAAAGGGCGGCACAAAGCCCTCTCTACCTGC AGCTCCCACCTCACGGTGGTTGTATTGTTCTTTGTCCCCTGTATTTTCTTGTACATGAGGCCTGTGGTCACTCACCCCAT AGACAAGGCAATGGCTGTGTCAGACTCAATCATCACACCCATGTTAAATCCCTTGATCTATACACTGAGGAATGCAGAGG TGAAAAGTGCCATGAAGAAACTCTGGATGAAATGGGAGGCTTTGGCTGGGAAATAACTGCAATGCTGA

The encoded GPCR10b protein is presented in Table 10D. The disclosed protein is 309 amino acids long and is denoted by SEQ ID NO:37. GPCR10b differs from GPCR10a by one amino acid residue: I133>T. Like GPCR10a, the Psort profile for GPCR10b predicts that this

sequence has a signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.6000. The most likely cleavage site for a peptide is between amino acids 39 and 40, *i.e.*, between the amino acid sequence VLE-NL (shown as a slash in Table10D) based on the SignalP result.

Table 10D. Encoded GPCR10b protein sequence (SEQ ID NO:37)
MENQNNVTEFILLGLTENLELWKIFSAVFLVMYVATVLE/NLLIVVTIITSQSLRSPMYFFLTFLSLLD VMFSSVVAPKVIVDTLSKSTTISLKGCLTQLFVEHFFGGVGIILLTVMAYDRYVAICKPLHYTITMSPR VCCLMVGGAWVGGFMHAMIQLLFMYQIPFCGPNIIDHFICDLFQLLTLACTDTHILGLLVTLNSGMMCV AIFLILIASYTVILCSLKSYSSKGRHKALSTCSSHLTVVVLFFVPCIFLYMRPVVTHPIDKAMAVSDSI ITPMLNPLIYTLRNAEVKSAMKKLWMKWEALAGK

Table 10E presents the BLASTP results for GPCR10b.

Table 10E. BLASTP Results for GPCR10b
Score = 999 (351.7 bits), Expect = 1.2e-100, P = 1.2e-100 Identities = 182/302 (60%), Positives = 231/302 (76%) with PIR-ID:S29710 olfactory receptor OR18 - rat Score = 757 (266.5 bits), Expect = 5.2e-75, P = 5.2e-75 Identities = 144/298 (48%), Positives = 200/298 (67%) with ACC:O95013 WUGSC:H_DJ0855D21.1 PROTEIN - Homo sapiens (Human), 312 aa.
Score = 757 (266.5 bits), Expect = 5.2e-75, P = 5.2e-75 Identities = 144/298 (48%), Positives = 200/298 (67%) with ACC:O95013 WUGSC:H_DJ0855D21.1 PROTEIN - Homo sapiens (Human), 312 aa. Score = 667 (234.8 bits), Expect = 1.1e-64, P = 1.1e-64 Identities = 131/300 (43%), Positives = 194/300 (64%), Frame = +1 with ACC:O43749 OLFACTORY RECEPTOR - Homo sapiens (Human), 312 aa.

GPCR10c

Another nucleotide sequence resulted when GPCR10a (6-L-19-A) was subjected to an exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such suitable sequences were then employed as the forward and reverse primers in a PCR amplification based on a wide range of cDNA libraries.

These primers were then employed in PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach,

testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. The resulting amplicon was gel purified, cloned and sequenced to high redundancy to provide the sequence reported below, which is designated as Accession Number 6-L-19-A-da1, or GPCR10c.

The nucleotide sequence for GPCR10c (943 bp, SEQ ID NO:38) is presented in Table 10F. The GPCR10c nucleotide sequence differs from GPCR10a by having six fewer nucleotides at the 5' end and two nucleotide changes: (numbered with respect to GPCR10a) G466>A and C834>T.

Table 10F. GPCR10c Nucleotide Sequence (SEQ ID NO:38)

ATGGAAAATCAAAACAATGTGACTGAATTCATTCTTCTGGGTCTCACAGAGAACCTGGAGCTGTGGAAAA
TATTTTCTGCTGTGTTTCTTGTCATGTATGTAGCCACAGTGCTGGAAAATCTACTTATTGTGGTAACTAT
TATCACAAGTCAGAGTCTGAGGTCACCTATGTATTTTTTTCTTACCTTCTTGTCCCTTTTGGATGTCATG
TTCTCATCTGTCGTTGCCCCCAAGGTGATTGTAGACACCCTCTCCAAGAGCACTACCATCTCTCTCAAAG
GCTGCCTCACCCAGCTGTTTGTGGAGCATTTCTTTGGTGGTGTGGGGATCATCCTCCTCACTGTGATGGC
CTATGACCGCTACGTGGCCATCTGTAAGCCCCTGCACTACACGATCATCATGAGTCCACGGGTGTGCTGC
CTAATGGTAGGAGGGGCTTGGGTGGGGGGATTTATGCACACAATGATACAACTTCTCTTCATGTATCAAA
TACCCTTCTGTGGTCCTAATATCATAGATCACTTTATATGTGATTTGTTTCAGTTGTTGACACTTGCCTG
CACGGACACCCACATCCTGGGCCTCTTAGTTACCCTCAACAGTGGGATGATGTGTGTGGCCATCTTTCTT
ATCTTAATTGCGTCCTACACGGTCATCCTATGCTCCCTGAAGTCTTACAGCTCTAAAGGGCGGCACAAAG
CCCTCTCTACCTGCAGCTCCCACCTCACGGTGGTTGTATTGTTCTTTGTCCCCTGTATTTTCTTGTACAT
GAGGCCTGTGGTCACTCACCCCATAGACAAGGCAATGGCTGTGTCAGACTCAATCATTACACCCATGTTA
AATCCCTTGATCTATACACTGAGGAATGCAGAGGTGAAAAGTGCCATGAAGAAACTCTGGATGAAATGGG
AGGCTTTGGCTGGGAAATAACTGCAATGCTGA

The coding region of GPCR10c is from nucleotide 1 to 928, giving the encoded GPCR10c protein, as presented in Table 10G. The disclosed protein is 309 amino acids long and is denoted by SEQ ID NO:83. GPCR10c differs from GPCR10a by one amino acid residue: A154>T. Like GPCR10a, the Psort profile for GPCR10c predicts that this sequence has a signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.6000. The most likely cleavage site for a peptide is between amino acids 39 and 40, *i.e.*, at the slash in the amino acid sequence VLE-NL (shown as a slash in Table10G) based on the SignalP result.

Table 10G. Encoded GPCR10c protein sequence (SEQ ID NO:83)
MENQNNVTEFILLGLTENLELWKIFSAVFLVMYVATVLE/NLLIVVTIITSQSLRSPMYFFLTFLSLLDVM FSSVVAPKVIVDTLSKSTTISLKGCLTQLFVEHFFGGVGIILLTVMAYDRYVAICKPLHYTIIMSPRVCC LMVGGAWVGGFMHTMIQLLFMYQIPFCGPNIIDHFICDLFQLLTLACTDTHILGLLVTLNSGMMCVAIFL ILIASYTVILCSLKSYSSKGRHKALSTCSSHLTVVVLFFVPCIFLYMRPVVTHPIDKAMAVSDSIITPML NPLIYTLRNAEVKSAMKKLWMKWEALAGK

Possible SNPs found for GPCR10 are listed in Table10H.

Table 10H: SNPs		
Base Position	**Base Before**	**Base After**
65	T	A(2)
120	T	Gap(2)
147	T	C(2)
234	A	G(3)
412	T	C(7)
471	G	A(2)
814	A	G(3)

Patp results include those listed in Table 10I.

Table 10I. Patp alignments of GPCR10

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob. P(N)
patp:Y90872 Human G protein-coupled receptor GTAR14-1 . ..	+1	796	2.0e-78
patp:Y90872 Human G protein-coupled receptor GTAR14-1 . ..	+1	796	2.0e-78
patp:Y92364 G protein-coupled receptor protein 4 H. sapiens.	+1	788	1.4e-77
patp:Y90874 Human G protein-coupled receptor GTAR14-5 .. .	+1	703	1.4e-68
patp:Y90874 Human G protein-coupled receptor GTAR14-5 .. .	+1	703	1.4e-68
patp:R27868 Odorant receptor clone F5 - Rattus rattus,.. .	+1	683	1.9e-66
patp:Y90873 Human G protein-coupled receptor GTAR14-3 ...	+1	671	3.5e-65

For example, a BLAST against Y90872, a 313 amino acid G-protein coupled receptor protein (GTAR14-1) from *Homo sapiens*, produced 148/297 (49%) identity, and 206/297 (69%) positives (E = 2.0e-78), with long segments of amino acid identity, as shown in Table 10J. WO 00/21999.

Table 10J. Alignment of GPCR10 with Y90872 (SEQ ID NO:89)

```
Length = 313  Plus Strand HSPs:
Score = 796 (280.2 bits), Expect = 2.0e-78, P = 2.0e-78
Identities = 148/297 (49%), Positives = 206/297 (69%), Frame = +1

GPCR10     3 NQNNVTEFILLGLTENLELWKIFSAVFLVMYVATVLENLLIVVTIITSQSLRSPMYFFLT 62
Y90872:    5 NQTRVTEFVFLGLTDNRVLEMLFFMAFSAIYMLTLSGNILIIIATVFTPSLHTPMYFFLS 64

GPCR10:   63 FLSLLDVMFSSVVAPKVIVDTLSKSTTISLKGCLTQLFVEHFFGGVGIILLTVMAYDRYV 122
Y90872:   65 NLSFIDICHSSVTVPKMLEGLLLERKTISFDNCITQLFFLHLFACAEIFLLIIVAYDRYV 124

GPCR10:  123 AICKPLHYTIIMSPRVCCLMVGGAWVGGFMHAMIQLLFMYQIPFCGPNIIDHFICDLFQL 182
Y90872:  125 AICTPLHYPNVMNMRVCIQLVFALWLGGTVHSLGQTFLTIRLPYCGPNIIDSYFCDVPLV 184

GPCR10:  183 LTLACTDTHILGLLVTLNSGMMCVAIFLILIASYTVILCSLKSYSSKGRHKALSTCSSHL 242
Y90872:  185 IKLACTDTYLTGILIVTNSGTISLSCFLAVVTSYMVILVSLRKHSAEGRQKALSTCSAHF 244

GPCR10:  243 TVVVLFFVPCIFLYMRPVVTHPIDKAMAVSDSIITPMLNPLIYTLRNAEVKSAMKKL 299
Y90872:  245 MVVALFFGPCIFIYTRPDTSFSIDKVVSVFYTVVTPLLNPFIYTLRNEEVKSAMKQL 301
```

The disclosed GPCR10 protein (SEQ ID NO:35) has good identity with a number of olfactory receptor proteins. The identity information used for ClustalW analysis is presented in Table 10K. Unless specifically addressed as GPCR10a GPCR10b, or GPCR10c, any reference to GPCR10 is assumed to encompass all variants. Residue differences between any GPCRX variant sequences herein are written to show the residue in the "a" variant and the residue position with respect to the "a" variant. GPCR residues in all following sequence alignments that differ between the individual GPCR variants are highlighted with a box and marked with the (o) symbol above the variant residue in all alignments herein. For example, the protein shown in line 1 of Table 10L depicts the sequence for GPCR10a, and the positions where GPCR10b or GPCR10c differs are marked with a (o) symbol and are highlighted with a box. All GPCR10 proteins have significant homology to olfactory receptor (OR) proteins:

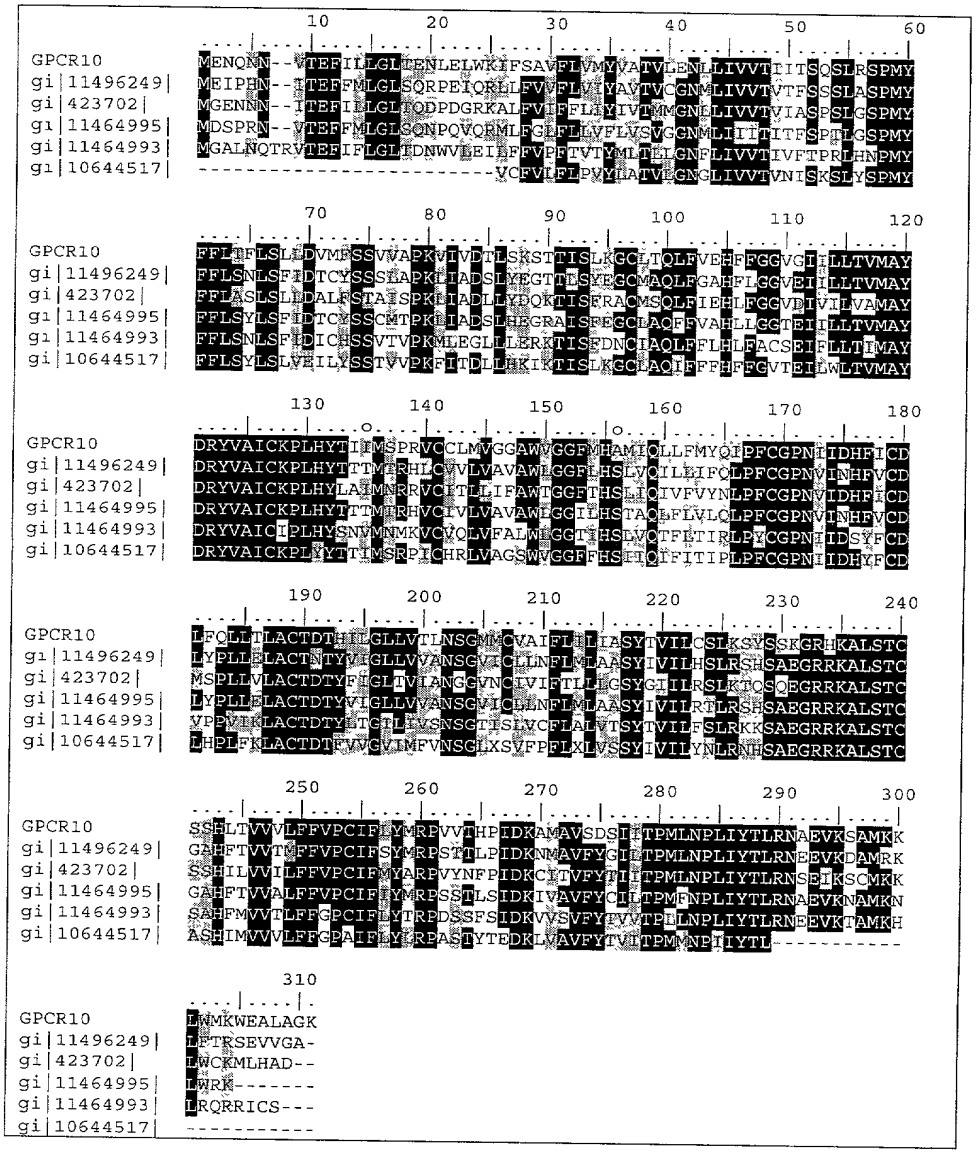
Table 10K. BLAST results for GPCR10

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
Gi|11496249|ref|NP_067343.1| (AB030895)	Odorant receptor MOR18 *Mus musculus*	308	184/306 (60%)	241/306 (78%)	6e-91
Gi|423702|pir||S29710	OR OR18 - rat	307	183/302 (60%)	232/302 (76%)	2e-88
Gi|11464995|ref|NP_065261.1| AB030896)	Odorant receptor A16 *Mus musculus*	302	175/302 (57%)	232/302 (75%)	8e-86
Gi|11464993|ref|NP_065260.1| (AB030894)	Odorant receptor MOR83 *Mus musculus*	308	157/297 (52%)	208/297 (69%)	3e-72
Gi|10644517|gb|AAG21323.1|AF271050_1 (AF271050)	Odorant receptor *Rattus norvegicus*	264	155/260 (59%)	202/260 (77%)	2e-71

This information is presented graphically in the multiple sequence alignment given in Table 10L (with GPCR10 being shown on line 1) as a ClustalW analysis comparing GPCR10 with related protein sequences.

Table 10L Information for the ClustalW proteins:

1) GPCR10 (SEQ ID NO:35)
2) gi|11496249|ref|NP_067343.1| odorant receptor 16 Mus musculus (SEQ ID NO:78)
3) gi|423702|pir||S29710 olfactory receptor OR18 - rat (SEQ ID NO:80)
4) gi|11464995|ref|NP_065261.1| gene for odorant receptor A16 Mus musculus (SEQ ID NO:79)
5) gi|11464993|ref|NP_065260.1| gene for odorant receptor MOR83 Mus musculus (SEQ ID NO:81)
6) gi|10644517|gb|AAG21323.1|AF271050_1 odorant receptor Rattus norvegicus (SEQ ID NO:82)



The presence of identifiable domains in GPCR10 was determined by searches using algorithms such as PROSITE, Blocks, Pfam, ProDomain, Prints and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (http:www.ebi.ac.uk/interpro/).

DOMAIN results for GPCR10 were collected from the Conserved Domain Database (CDD) with Reverse Position Specific BLAST. This BLAST samples domains found in the Smart and Pfam collections. The results are listed in Table 10M with the statistics and domain description. The results indicate that this protein contains the following protein domains (as defined by Interpro) at the indicated positions: domain name 7tm_1 (InterPro) 7 transmembrane receptor (rhodopsin family) at amino acid positions 39 to 213. This indicates that the sequence of GPCR10 has properties similar to those of other proteins known to contain this domain and similar to the properties of this domain.

Table 10M. DOMAIN results for GPCR10

gnl|Pfam|pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family) Length = 377 (SEQ ID NO:45).

Score = 93.6 bits (231), Expect = 1e-20

```
GPCR10  ENLLIVVTIITSQSLRSPMYFFLTFLSLLDVMFSSVVAPKVIVDTLSKSTTISLKGCLTQ
TM7     GNVLVCMAVSREKALQTTTNYLIVSLAVADLLVATLVMPWVVYLEVVGEWKFSRIHCDIF

GPCR10  LFVEHFFGGVGIILLLTVMAYDRYVAICKPLHYTIIM-SPRVCCLMVGGAWVGGFMHAMIQ
TM7     VTLDVMMCTASILNLCAISIDRYTAVAMPMLYNTRYSSKRRVTVMIAIVWVLSFTISCPM

GPCR10  LLF-MYQIPFCG--------PNIIDHFICDLFQL---LTLACTDTHILGLLVTLNSGMMC
TM7     LFGLNNTDQNE--------CIIANP--------AFVVYSSIVSFYV----PFIVTLLVYI

GPCR10  VAIFLILI----------------------------------------------------
TM7     KIYIVLRRRRKRVNTKRSSRAFRANLKAPLKGNCTHPEDMKLCTVIMKSNGSFPVNRRRV
```

The similarity information for the GPCR10 protein and nucleic acid disclosed herein suggest that GPCR10 may have important structural and/or physiological functions characteristic of the Olfactory Receptor family and the GPCR family. Therefore, the nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene

ablation), and (v) a composition promoting tissue regeneration in vitro and in vivo (vi) biological defense weapon. The novel nucleic acid encoding GPCR10, and the GPCR10 protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

The nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in used in the treatment of infections such as bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease; multiple sclerosis; and treatment of Albright hereditary ostoeodystrophy, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy, and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome and/or other pathologies and disorders. The polypeptides can be used as immunogens to produce antibodies specific for the invention, and as vaccines. They can also be used to screen for potential agonist and antagonist compounds. For example, a cDNA encoding the GPCR-like protein may be useful in gene therapy, and the GPCR-like protein may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease; multiple sclerosis; and treatment of Albright hereditary ostoeodystrophy, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy, and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome and/or other pathologies and disorders. These materials are further useful in the generation of antibodies that bind immuno-specifically to the novel GPCR10 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-GPCRX Antibodies" section below. For example the disclosed GPCR10 protein has multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated GPCR10 epitope is from about

amino acids 5 to 15. In another embodiment, a GPCR10 epitope is from about amino acids 225 to 240. In additional embodiments, GPCR10 epitopes are from amino acids 260 to 270 and from amino acids 290 to 310. These novel proteins can also be used to develop assay system for functional analysis.

A summary of the GPCRX nucleic acids and proteins of the invention is provided in Table 11.

TABLE 11: Summary Of Nucleic Acids And Proteins Of The Invention

Name	Tables	Clone; Description of Homolog	Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO
GPCR1	1A, 1B, 1E, 1F, 1I, 1J	GPCR1a: ba113a10_B, olfactory receptor	1	2
		GPCR1b: ba32713_A, olfactory receptor	3	4
		GPCR1c: ba113a10_C, olfactory receptor	5	6
GPCR2	2A, 2B	GPCR2a: 11612531_1, olfactory receptor	7	8
	2I, 2J	GPCR2b: 11612531_1_da1, olfactory receptor	84	85
GPCR3	3A, 3B	GPCR3: ba145L22_B, olfactory receptor	9	10
GPCR4	4A, 4B, 4C, 4G, 4H	GPCR4a1: dj408b20_C, olfactory receptor	11	12
		GPCR4a2: dj408b20_C_da1, olfactory receptor	13	
		GPCR4a3: CG55358_03, olfactory receptor	16	17
GPCR5	5A, 5B, 5C, 5D 5G, 5H	GPCR5a1: 115-a-12-A, olfactory receptor	18	19
		GPCR5a2: 115-a-12-B, olfactory receptor	20	21
		GPCR5a3: 115-a-12-A_da1, olfactory receptor	22	23
GPCR6	6A, 6B	GPCR6: 6-L-19-C, olfactory receptor	24	25
GPCR7	7A, 7B	GPCR7: dj313i6_D olfactory receptor	28	29
GPCR8	8A, 8B	GPCR8: dj408b20_A, olfactory receptor	30	31
GPCR9	9A, 9B	GPCR9: 6-L-19-B, olfactory receptor	32	33
GPCR10	10A, 10B, 10C, 10D, 10F, 10G	GPCR10a: 6-L-19-A, olfactory receptor	34	35
		GPCR10b: 6-L-19-A1, olfactory receptor	36	37
		GPCR10c: 6-L-19-A_da1, olfactory receptor	38	83

GPCRX Nucleic Acids and Polypeptides

One aspect of the invention pertains to isolated nucleic acid molecules that encode GPCRX polypeptides or biologically active portions thereof. Also included in the invention are nucleic acid fragments sufficient for use as hybridization probes to identify GPCRX-encoding nucleic acids (*e.g.*, GPCRX mRNAs) and fragments for use as PCR primers for the amplification and/or mutation of GPCRX nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and

derivatives, fragments and homologs thereof. The nucleic acid molecule may be single-stranded or double-stranded, but preferably is comprised double-stranded DNA.

An GPCRX nucleic acid can encode a mature GPCRX polypeptide. As used herein, a "mature" form of a polypeptide or protein disclosed in the present invention is the product of a naturally occurring polypeptide or precursor form or proprotein. The naturally occurring polypeptide, precursor or proprotein includes, by way of nonlimiting example, the full-length gene product, encoded by the corresponding gene. Alternatively, it may be defined as the polypeptide, precursor or proprotein encoded by an ORF described herein. The product "mature" form arises, again by way of nonlimiting example, as a result of one or more naturally occurring processing steps as they may take place within the cell, or host cell, in which the gene product arises. Examples of such processing steps leading to a "mature" form of a polypeptide or protein include the cleavage of the N-terminal methionine residue encoded by the initiation codon of an ORF, or the proteolytic cleavage of a signal peptide or leader sequence. Thus a mature form arising from a precursor polypeptide or protein that has residues 1 to N, where residue 1 is the N-terminal methionine, would have residues 2 through N remaining after removal of the N-terminal methionine. Alternatively, a mature form arising from a precursor polypeptide or protein having residues 1 to N, in which an N-terminal signal sequence from residue 1 to residue M is cleaved, would have the residues from residue M+1 to residue N remaining. Further as used herein, a "mature" form of a polypeptide or protein may arise from a step of post-translational modification other than a proteolytic cleavage event. Such additional processes include, by way of non-limiting example, glycosylation, myristoylation or phosphorylation. In general, a mature polypeptide or protein may result from the operation of only one of these processes, or a combination of any of them.

The term "probes", as utilized herein, refers to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), 100 nt, or as many as approximately, *e.g.*, 6,000 nt, depending upon the specific use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are generally obtained from a natural or recombinant source, are highly specific, and much slower to hybridize than shorter-length oligomer probes. Probes may be single- or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

The term "isolated" nucleic acid molecule, as utilized herein, is one, which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5'- and 3'-termini of the nucleic acid) in the genomic DNA of the

organism from which the nucleic acid is derived. For example, in various embodiments, the isolated GPCRX nucleic acid molecules can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell/tissue from which the nucleic acid is derived (*e.g.*, brain, heart, liver, spleen, etc.). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84, or a complement of this aforementioned nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84 as a hybridization probe, GPCRX molecules can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook, *et al.*, (eds.), MOLECULAR CLONING: A LABORATORY MANUAL 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to GPCRX nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

As used herein, the term "oligonucleotide" refers to a series of linked nucleotide residues, which oligonucleotide has a sufficient number of nucleotide bases to be used in a PCR reaction. A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue. Oligonucleotides comprise portions of a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment of the invention, an oligonucleotide comprising a nucleic acid molecule less than 100 nt in length would further comprise at least 6 contiguous nucleotides of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84, or a complement thereof. Oligonucleotides may be chemically synthesized and may also be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84, or a portion of this nucleotide sequence (*e.g.*, a fragment that can be used as a probe or primer or a fragment encoding a biologically-active portion of an GPCRX polypeptide). A nucleic acid molecule that is complementary to the nucleotide sequence shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84 is one that is sufficiently complementary to the nucleotide sequence shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84 that it can hydrogen bond with little or no mismatches to the nucleotide sequence shown SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84, thereby forming a stable duplex.

As used herein, the term "complementary" refers to Watson-Crick or Hoogsteen base pairing between nucleotides units of a nucleic acid molecule, and the term "binding" means the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, van der Waals, hydrophobic interactions, and the like. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

Fragments provided herein are defined as sequences of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, respectively, and are at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice. Derivatives are nucleic acid sequences or amino acid sequences formed from the native compounds either directly or by modification or partial substitution. Analogs are nucleic acid sequences or amino acid sequences that have a structure similar to, but not identical to, the native compound but differs from it in respect to certain components or side chains. Analogs may be synthetic or from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type. Homologs are nucleic acid sequences or amino acid sequences of a particular gene that are derived from different species.

Derivatives and analogs may be full length or other than full length, if the derivative or analog contains a modified nucleic acid or amino acid, as described below. Derivatives or

analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the invention, in various embodiments, by at least about 70%, 80%, or 95% identity (with a preferred identity of 80-95%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the aforementioned proteins under stringent, moderately stringent, or low stringent conditions. See *e.g.* Ausubel, *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below.

A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences encode those sequences coding for isoforms of GPCRX polypeptides. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the invention, homologous nucleotide sequences include nucleotide sequences encoding for an GPCRX polypeptide of species other than humans, including, but not limited to: vertebrates, and thus can include, *e.g.*, frog, mouse, rat, rabbit, dog, cat cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the exact nucleotide sequence encoding human GPCRX protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84, as well as a polypeptide possessing GPCRX biological activity. Various biological activities of the GPCRX proteins are described below.

An GPCRX polypeptide is encoded by the open reading frame ("ORF") of an GPCRX nucleic acid. An ORF corresponds to a nucleotide sequence that could potentially be translated into a polypeptide. A stretch of nucleic acids comprising an ORF is uninterrupted by a stop codon. An ORF that represents the coding sequence for a full protein begins with an ATG "start" codon and terminates with one of the three "stop" codons, namely, TAA, TAG, or TGA. For the purposes of this invention, an ORF may be any part of a coding sequence, with or without a start codon, a stop codon, or both. For an ORF to be considered as a good candidate for coding for a *bona fide* cellular protein, a minimum size requirement is often set, *e.g.*, a stretch of DNA that would encode a protein of 50 amino acids or more.

The nucleotide sequences determined from the cloning of the human GPCRX genes allows for the generation of probes and primers designed for use in identifying and/or cloning GPCRX homologues in other cell types, *e.g.* from other tissues, as well as GPCRX homologues from other vertebrates. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 consecutive sense strand nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84; or an anti-sense strand nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84; or of a naturally occurring mutant of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84.

Probes based on the human GPCRX nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe further comprises a label group attached thereto, *e.g.* the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissues which mis-express an GPCRX protein, such as by measuring a level of an GPCRX-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting GPCRX mRNA levels or determining whether a genomic GPCRX gene has been mutated or deleted.

"A polypeptide having a biologically-active portion of an GPCRX polypeptide" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically-active portion of GPCRX" can be prepared by isolating a portion SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84 that encodes a polypeptide having an GPCRX biological activity (the biological activities of the GPCRX proteins are described below), expressing the encoded portion of GPCRX protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of GPCRX.

GPCRX Nucleic Acid and Polypeptide Variants

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences shown SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84 due to degeneracy of the genetic code and thus encode the same GPCRX proteins as that encoded by the nucleotide sequences shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22,

24, 28, 30, 32, 34, 36, 38, and 84. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NOS:2, 4, 6, 8, 10, 12, 17, 19, 21, 23, 25, 29, 31, 33, 35, 37, 83, and 85.

In addition to the human GPCRX nucleotide sequences shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84 it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the GPCRX polypeptides may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the GPCRX genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame (ORF) encoding an GPCRX protein, preferably a vertebrate GPCRX protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the GPCRX genes. Any and all such nucleotide variations and resulting amino acid polymorphisms in the GPCRX polypeptides, which are the result of natural allelic variation and that do not alter the functional activity of the GPCRX polypeptides, are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding GPCRX proteins from other species, and thus that have a nucleotide sequence that differs from the human sequence SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84 are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the GPCRX cDNAs of the invention can be isolated based on their homology to the human GPCRX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500, 750, 1000, 1500, or 2000 or more nucleotides in length. In yet another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other.

Homologs (*i.e.*, nucleic acids encoding GPCRX proteins derived from species other than human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high

stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5 °C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at Tm, 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes, primers or oligonucleotides (*e.g.*, 10 nt to 50 nt) and at least about 60°C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions are hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C, followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequences of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84 corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84 or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency

hybridization conditions are hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55°C, followed by one or more washes in 1X SSC, 0.1% SDS at 37°C. Other conditions of moderate stringency that may be used are well-known within the art. See, *e.g.*, Ausubel, et *al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990; GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequences of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84 or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that may be used are well known in the art (*e.g.*, as employed for cross-species hybridizations). See, *e.g.*, Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981. *Proc Natl Acad Sci USA* **78**: 6789-6792.

Conservative Mutations

In addition to naturally-occurring allelic variants of GPCRX sequences that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequences of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84 thereby leading to changes in the amino acid sequences of the encoded GPCRX proteins, without altering the functional ability of said GPCRX proteins. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of SEQ ID NOS:2, 4, 6, 8, 10, 12, 17, 19, 21, 23, 25, 29, 31, 33, 35, 37, 83, and 85. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequences of the GPCRX proteins without altering their biological activity, whereas an "essential" amino acid residue is required for such biological activity. For example, amino acid residues that are conserved among the GPCRX proteins of the invention are predicted to be particularly non-amenable to alteration. Amino acids for which conservative substitutions can be made are well-known within the art.

Another aspect of the invention pertains to nucleic acid molecules encoding GPCRX proteins that contain changes in amino acid residues that are not essential for activity. Such GPCRX proteins differ in amino acid sequence from SEQ ID NOS:2, 4, 6, 8, 10, 12, 17, 19, 21, 23, 25, 29, 31, 33, 35, 37, 83, and 85 yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 45% homologous to the amino acid sequences of SEQ ID NOS:2, 4, 6, 8, 10, 12, 17, 19, 21, 23, 25, 29, 31, 33, 35, 37, 83, and 85. Preferably, the protein encoded by the nucleic acid molecule is at least about 60% homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 17, 19, 21, 23, 25, 29, 31, 33, 35, 37, 83, and 85; more preferably at least about 70% homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 17, 19, 21, 23, 25, 29, 31, 33, 35, 37, 83, and 85; still more preferably at least about 80% homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 17, 19, 21, 23, 25, 29, 31, 33, 35, 37, 83, and 85; even more preferably at least about 90% homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 17, 19, 21, 23, 25, 29, 31, 33, 35, 37, 83, and 85; and most preferably at least about 95% homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 17, 19, 21, 23, 25, 29, 31, 33, 35, 37, 83, and 85.

An isolated nucleic acid molecule encoding an GPCRX protein homologous to the protein of SEQ ID NOS:2, 4, 6, 8, 10, 12, 17, 19, 21, 23, 25, 29, 31, 33, 35, 37, 83, and 85 can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84 such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced into SEQ ID NOS:2, 4, 6, 8, 10, 12, 17, 19, 21, 23, 25, 29, 31, 33, 35, 37, 83, and 85 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted, non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined within the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted non-essential amino acid residue in the GPCRX protein is replaced with another amino acid residue from the same side

chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of an GPCRX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for GPCRX biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84, the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

The relatedness of amino acid families may also be determined based on side chain interactions. Substituted amino acids may be fully conserved "strong" residues or fully conserved "weak" residues. The "strong" group of conserved amino acid residues may be any one of the following groups: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW, wherein the single letter amino acid codes are grouped by those amino acids that may be substituted for each other. Likewise, the "weak" group of conserved residues may be any one of the following: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, VLIM, HFY, wherein the letters within each group represent the single letter amino acid code.

In one embodiment, a mutant GPCRX protein can be assayed for (*i*) the ability to form protein:protein interactions with other GPCRX proteins, other cell-surface proteins, or biologically-active portions thereof, (*ii*) complex formation between a mutant GPCRX protein and an GPCRX ligand; or (*iii*) the ability of a mutant GPCRX protein to bind to an intracellular target protein or biologically-active portion thereof; (*e.g.* avidin proteins).

In yet another embodiment, a mutant GPCRX protein can be assayed for the ability to regulate a specific biological function (*e.g.*, regulation of insulin release).

Antisense Nucleic Acids

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein (*e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence). In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire GPCRX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of an GPCRX protein of SEQ ID NOS:2, 4, 6, 8, 10, 12, 17, 19, 21, 23, 25, 29, 31, 33, 35, 37, 83, and 85, or antisense nucleic acids

complementary to an GPCRX nucleic acid sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84, are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding an GPCRX protein. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding the GPCRX protein. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding the GPCRX protein disclosed herein, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of GPCRX mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of GPCRX mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of GPCRX mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally-occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids (*e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used).

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil,

uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding an GPCRX protein to thereby inhibit expression of the protein (*e.g.*, by inhibiting transcription and/or translation). The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface (*e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens). The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient nucleic acid molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other. See, *e.g.*, Gaultier, *et al.*, 1987. *Nucl. Acids Res.* **15**: 6625-6641. The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (see, *e.g.*, Inoue, *et al.* 1987. *Nucl. Acids Res.* **15**: 6131-6148) or a chimeric RNA-DNA analogue (see, *e.g.*, Inoue, *et al.*, 1987. *FEBS Lett.* **215**: 327-330.

Ribozymes and PNA Moieties

Nucleic acid modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified

nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

In one embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes as described in Haselhoff and Gerlach 1988. *Nature* 334: 585-591) can be used to catalytically cleave GPCRX mRNA transcripts to thereby inhibit translation of GPCRX mRNA. A ribozyme having specificity for an GPCRX-encoding nucleic acid can be designed based upon the nucleotide sequence of an GPCRX cDNA disclosed herein (*i.e.*, SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in an GPCRX-encoding mRNA. See, *e.g.*, U.S. Patent 4,987,071 to Cech, *et al.* and U.S. Patent 5,116,742 to Cech, *et al.* GPCRX mRNA can also be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, GPCRX gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the GPCRX nucleic acid (*e.g.*, the GPCRX promoter and/or enhancers) to form triple helical structures that prevent transcription of the GPCRX gene in target cells. *See, e.g.,* Helene, 1991. *Anticancer Drug Des.* 6: 569-84; Helene, *et al.* 1992. *Ann. N.Y. Acad. Sci.* 660: 27-36; Maher, 1992. *Bioassays* 14: 807-15.

In various embodiments, the GPCRX nucleic acids can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids. *See, e.g.,* Hyrup, *et al.*, 1996. *Bioorg Med Chem* 4: 5-23. As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics (*e.g.*, DNA mimics) in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup, *et al.*, 1996. *supra*; Perry-O'Keefe, *et al.*, 1996. *Proc. Natl. Acad. Sci. USA* 93: 14670-14675.

PNAs of GPCRX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene

expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of GPCRX can also be used, for example, in the analysis of single base pair mutations in a gene (*e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S_1 nucleases (*see,* Hyrup, *et al.,* 1996.*supra*); or as probes or primers for DNA sequence and hybridization (*see,* Hyrup, *et al.*, 1996, *supra*; Perry-O'Keefe, *et al.,* 1996. *supra*).

In another embodiment, PNAs of GPCRX can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of GPCRX can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes (*e.g.*, RNase H and DNA polymerases) to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (*see,* Hyrup, et al., 1996. *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup, *et al.,* 1996. *supra* and Finn, *et al.*, 1996. *Nucl Acids Res* 24: 3357-3363. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA. *See, e.g.,* Mag, *et al.*, 1989. *Nucl Acid Res* 17: 5973-5988. PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment. *See, e.g.,* Finn, *et al.*, 1996. *supra.* Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. *See, e.g.,* Petersen, *et al.*, 1975. *Bioorg. Med. Chem. Lett.* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (*see, e.g.*, Letsinger, *et al.*, 1989. *Proc. Natl. Acad. Sci. U.S.A.* 86: 6553-6556; Lemaitre, *et al.*, 1987. *Proc. Natl. Acad. Sci.* 84: 648-652; PCT Publication No. WO88/09810) or the blood-brain barrier (*see, e.g.*, PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (*see, e.g.,* Krol, *et al.*, 1988. *BioTechniques* 6:958-976) or intercalating agents (*see, e.g.*, Zon, 1988. *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, and the like.

GPCRX Polypeptides

A polypeptide according to the invention includes a polypeptide including the amino acid sequence of GPCRX polypeptides whose sequences are provided in SEQ ID NOS:2, 4, 6, 8, 10, 12, 17, 19, 21, 23, 25, 29, 31, 33, 35, 37, 83, and 85. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residues shown in SEQ ID NOS:2, 4, 6, 8, 10, 12, 17, 19, 21, 23, 25, 29, 31, 33, 35, 37, 83, and 85 while still encoding a protein that maintains its GPCRX activities and physiological functions, or a functional fragment thereof.

In general, an GPCRX variant that preserves GPCRX-like function includes any variant in which residues at a particular position in the sequence have been substituted by other amino acids, and further include the possibility of inserting an additional residue or residues between two residues of the parent protein as well as the possibility of deleting one or more residues from the parent sequence. Any amino acid substitution, insertion, or deletion is encompassed by the invention. In favorable circumstances, the substitution is a conservative substitution as defined above.

One aspect of the invention pertains to isolated GPCRX proteins, and biologically-active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also provided are polypeptide fragments suitable for use as immunogens to raise anti-GPCRX antibodies. In one embodiment, native GPCRX proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, GPCRX proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, an GPCRX protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" polypeptide or protein or biologically-active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the GPCRX protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of GPCRX proteins in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly-produced. In one embodiment, the language "substantially free of cellular material" includes preparations of GPCRX proteins having less than about 30% (by dry weight) of non-GPCRX proteins (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-GPCRX proteins, still more preferably less than about 10% of non-GPCRX proteins, and most preferably less than about 5% of non-GPCRX proteins. When the GPCRX protein or

biologically-active portion thereof is recombinantly-produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the GPCRX protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of GPCRX proteins in which the protein is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of GPCRX proteins having less than about 30% (by dry weight) of chemical precursors or non-GPCRX chemicals, more preferably less than about 20% chemical precursors or non-GPCRX chemicals, still more preferably less than about 10% chemical precursors or non-GPCRX chemicals, and most preferably less than about 5% chemical precursors or non-GPCRX chemicals.

Biologically-active portions of GPCRX proteins include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequences of the GPCRX proteins (*e.g.*, the amino acid sequence shown in SEQ ID NOS:2, 4, 6, 8, 10, 12, 17, 19, 21, 23, 25, 29, 31, 33, 35, 37, 83, and 85) that include fewer amino acids than the full-length GPCRX proteins, and exhibit at least one activity of an GPCRX protein. Typically, biologically-active portions comprise a domain or motif with at least one activity of the GPCRX protein. A biologically-active portion of an GPCRX protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acid residues in length.

Moreover, other biologically-active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native GPCRX protein.

In an embodiment, the GPCRX protein has an amino acid sequence shown in SEQ ID NOS:2, 4, 6, 8, 10, 12, 17, 19, 21, 23, 25, 29, 31, 33, 35, 37, 83, and 85. In other embodiments, the GPCRX protein is substantially homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 17, 19, 21, 23, 25, 29, 31, 33, 35, 37, 83, and 85, and retains the functional activity of the protein of SEQ ID NOS:2, 4, 6, 8, 10, 12, 17, 19, 21, 23, 25, 29, 31, 33, 35, 37, 83, and 85, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail, below. Accordingly, in another embodiment, the GPCRX protein is a protein that comprises an amino acid sequence at least about 45% homologous to the amino acid sequence SEQ ID NOS:2, 4, 6, 8, 10, 12, 17, 19, 21, 23, 25, 29, 31, 33, 35, 37, 83, and 85, and retains the functional activity of

the GPCRX proteins of SEQ ID NOS:2, 4, 6, 8, 10, 12, 17, 19, 21, 23, 25, 29, 31, 33, 35, 37, 83, and 85.

Determining Homology Between Two or More Sequences

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. *See,* Needleman and Wunsch, 1970. *J Mol Biol* 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84.

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region.

Chimeric and Fusion Proteins

The invention also provides GPCRX chimeric or fusion proteins. As used herein, an GPCRX "chimeric protein" or "fusion protein" comprises an GPCRX polypeptide operatively-linked to a non-GPCRX polypeptide. An "GPCRX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to an GPCRX protein (SEQ ID NOS:2, 4, 6, 8, 10, 12, 17, 19, 21, 23, 25, 29, 31, 33, 35, 37, 83, and 85), whereas a "non-GPCRX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the GPCRX protein, *e.g.*, a protein that is different from the GPCRX protein and that is derived from the same or a different organism. Within an GPCRX fusion protein the GPCRX polypeptide can correspond to all or a portion of an GPCRX protein. In one embodiment, an GPCRX fusion protein comprises at least one biologically-active portion of an GPCRX protein. In another embodiment, an GPCRX fusion protein comprises at least two biologically-active portions of an GPCRX protein. In yet another embodiment, an GPCRX fusion protein comprises at least three biologically-active portions of an GPCRX protein. Within the fusion protein, the term "operatively-linked" is intended to indicate that the GPCRX polypeptide and the non-GPCRX polypeptide are fused in-frame with one another. The non-GPCRX polypeptide can be fused to the N-terminus or C-terminus of the GPCRX polypeptide.

In one embodiment, the fusion protein is a GST-GPCRX fusion protein in which the GPCRX sequences are fused to the C-terminus of the GST (glutathione S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant GPCRX polypeptides.

In another embodiment, the fusion protein is an GPCRX protein containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of GPCRX can be increased through use of a heterologous signal sequence.

In yet another embodiment, the fusion protein is an GPCRX-immunoglobulin fusion protein in which the GPCRX sequences are fused to sequences derived from a member of the immunoglobulin protein family. The GPCRX-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between an GPCRX ligand and an GPCRX protein on the surface of a cell, to thereby suppress GPCRX-mediated signal transduction *in vivo*. The GPCRX-immunoglobulin fusion proteins can be used to affect the bioavailability of an GPCRX cognate ligand. Inhibition of the GPCRX ligand/GPCRX interaction may be useful therapeutically for both the treatment of

proliferative and differentiative disorders, as well as modulating (*e.g.* promoting or inhibiting) cell survival. Moreover, the GPCRX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-GPCRX antibodies in a subject, to purify GPCRX ligands, and in screening assays to identify molecules that inhibit the interaction of GPCRX with an GPCRX ligand.

An GPCRX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (*see, e.g.,* Ausubel, *et al.* (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). An GPCRX-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the GPCRX protein.

GPCRX Agonists and Antagonists

The invention also pertains to variants of the GPCRX proteins that function as either GPCRX agonists (*i.e.*, mimetics) or as GPCRX antagonists. Variants of the GPCRX protein can be generated by mutagenesis (*e.g.*, discrete point mutation or truncation of the GPCRX protein). An agonist of the GPCRX protein can retain substantially the same, or a subset of, the biological activities of the naturally occurring form of the GPCRX protein. An antagonist of the GPCRX protein can inhibit one or more of the activities of the naturally occurring form of the GPCRX protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the GPCRX protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the GPCRX proteins.

Variants of the GPCRX proteins that function as either GPCRX agonists (*i.e.*, mimetics) or as GPCRX antagonists can be identified by screening combinatorial libraries of mutants (*e.g.*, truncation mutants) of the GPCRX proteins for GPCRX protein agonist or antagonist activity. In one embodiment, a variegated library of GPCRX variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of GPCRX variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential GPCRX sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of GPCRX sequences therein. There are a variety of methods which can be used to produce libraries of potential GPCRX variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential GPCRX sequences. Methods for synthesizing degenerate oligonucleotides are well-known within the art. *See, e.g.*, Narang, 1983. *Tetrahedron* 39: 3; Itakura, *et al.*, 1984. *Annu. Rev. Biochem.* 53: 323; Itakura, *et al.*, 1984. *Science* 198: 1056; Ike, *et al.*, 1983. *Nucl. Acids Res.* 11: 477.

Polypeptide Libraries

In addition, libraries of fragments of the GPCRX protein coding sequences can be used to generate a variegated population of GPCRX fragments for screening and subsequent selection of variants of an GPCRX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of an GPCRX coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double-stranded DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S_1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, expression libraries can be derived which encodes N-terminal and internal fragments of various sizes of the GPCRX proteins.

Various techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of GPCRX proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene

libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify GPCRX variants. *See, e.g.,* Arkin and Yourvan, 1992. *Proc. Natl. Acad. Sci. USA* 89: 7811-7815; Delgrave, *et al.*, 1993. *Protein Engineering* 6:327-331.

Anti-GPCRX Antibodies

The invention encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the GPCRX polypeptides of said invention.

An isolated GPCRX protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind to GPCRX polypeptides using standard techniques for polyclonal and monoclonal antibody preparation. The full-length GPCRX proteins can be used or, alternatively, the invention provides antigenic peptide fragments of GPCRX proteins for use as immunogens. The antigenic GPCRX peptides comprises at least 4 amino acid residues of the amino acid sequence shown in SEQ ID NOS:2, 4, 6, 8, 10, 12, 17, 19, 21, 23, 25, 29, 31, 33, 35, 37, 83, and 85 and encompasses an epitope of GPCRX such that an antibody raised against the peptide forms a specific immune complex with GPCRX. Preferably, the antigenic peptide comprises at least 6, 8, 10, 15, 20, or 30 amino acid residues. Longer antigenic peptides are sometimes preferable over shorter antigenic peptides, depending on use and according to methods well known to someone skilled in the art.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a region of GPCRX that is located on the surface of the protein (*e.g.*, a hydrophilic region). As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation (*see, e.g.*, Hopp and Woods, 1981. *Proc. Nat. Acad. Sci. USA* 78: 3824-3828; Kyte and Doolittle, 1982. *J. Mol. Biol.* 157: 105-142, each incorporated herein by reference in their entirety).

As disclosed herein, GPCRX protein sequences of SEQ ID NOS:2, 4, 6, 8, 10, 12, 17, 19, 21, 23, 25, 29, 31, 33, 35, 37, 83, and 85, or derivatives, fragments, analogs or homologs thereof, may be utilized as immunogens in the generation of antibodies that immunospecifically-bind

these protein components. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically-active portions of immunoglobulin molecules, *i.e.*, molecules that contain an antigen binding site that specifically-binds (immunoreacts with) an antigen, such as GPCRX. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab} and $F_{(ab')2}$ fragments, and an F_{ab} expression library. In a specific embodiment, antibodies to human GPCRX proteins are disclosed. Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies to an GPCRX protein sequence of SEQ ID NOS:2, 4, 6, 8, 10, 12, 17, 19, 21, 23, 25, 29, 31, 33, 35, 37, 83, and 85, or a derivative, fragment, analog or homolog thereof. Some of these proteins are discussed below.

For the production of polyclonal antibodies, various suitable host animals (*e.g.*, rabbit, goat, mouse or other mammal) may be immunized by injection with the native protein, or a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, recombinantly-expressed GPCRX protein or a chemically-synthesized GPCRX polypeptide. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (*e.g.*, aluminum hydroxide), surface active substances (*e.g.*, lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), human adjuvants such as *Bacille Calmette-Guerin* and *Corynebacterium parvum*, or similar immunostimulatory agents. If desired, the antibody molecules directed against GPCRX can be isolated from the mammal (*e.g.*, from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction.

The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of GPCRX. A monoclonal antibody composition thus typically displays a single binding affinity for a particular GPCRX protein with which it immunoreacts. For preparation of monoclonal antibodies directed towards a particular GPCRX protein, or derivatives, fragments, analogs or homologs thereof, any technique that provides for the production of antibody molecules by continuous cell line culture may be utilized. Such techniques include, but are not limited to, the hybridoma technique (*see, e.g.,* Kohler & Milstein, 1975. *Nature* 256: 495-497); the trioma technique; the human B-cell hybridoma technique (*see, e.g.,* Kozbor, *et al.,* 1983. *Immunol. Today* 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (*see, e.g.,* Cole, *et al.,* 1985. In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the invention and may be produced by

using human hybridomas (*see, e.g.,* Cote, *et al.,* 1983. *Proc Natl Acad Sci USA* 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus *in vitro* (*see, e.g.,* Cole, *et al.,* 1985. In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Each of the above citations is incorporated herein by reference in their entirety.

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an GPCRX protein (*see, e.g.,* U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (*see, e.g.,* Huse, *et al.,* 1989. *Science* 246: 1275-1281) to allow rapid and effective identification of monoclonal F_{ab} fragments with the desired specificity for an GPCRX protein or derivatives, fragments, analogs or homologs thereof. Non-human antibodies can be "humanized" by techniques well known in the art. *See, e.g.,* U.S. Patent No. 5,225,539. Antibody fragments that contain the idiotypes to an GPCRX protein may be produced by techniques known in the art including, but not limited to: (*i*) an $F_{(ab')2}$ fragment produced by pepsin digestion of an antibody molecule; (*ii*) an F_{ab} fragment generated by reducing the disulfide bridges of an $F_{(ab')2}$ fragment; (*iii*) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing agent; and (*iv*) F_v fragments.

Additionally, recombinant anti-GPCRX antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in International Application No. PCT/US86/02269; European Patent Application No. 184,187; European Patent Application No. 171,496; European Patent Application No. 173,494; PCT International Publication No. WO 86/01533; U.S. Patent No. 4,816,567; U.S. Pat. No. 5,225,539; European Patent Application No. 125,023; Better, *et al.,* 1988. *Science* 240: 1041-1043; Liu, *et al.,* 1987. *Proc. Natl. Acad. Sci. USA* 84: 3439-3443; Liu, *et al.,* 1987. *J. Immunol.* 139: 3521-3526; Sun, *et al.,* 1987. *Proc. Natl. Acad. Sci. USA* 84: 214-218; Nishimura, *et al.,* 1987. *Cancer Res.* 47: 999-1005; Wood, *et al.,* 1985. *Nature* 314 :446-449; Shaw, *et al.,* 1988. *J. Natl. Cancer Inst.* 80: 1553-1559); Morrison(1985) *Science* 229:1202-1207; Oi, *et al.* (1986) *BioTechniques* 4:214; Jones, *et al.,* 1986. *Nature* 321: 552-525; Verhoeyan, *et al.,* 1988. *Science* 239: 1534; and Beidler, *et al.,* 1988. *J. Immunol.* 141: 4053-4060. Each of the above citations are incorporated herein by reference in their entirety.

In one embodiment, methods for the screening of antibodies that possess the desired specificity include, but are not limited to, enzyme-linked immunosorbent assay (ELISA) and

other immunologically-mediated techniques known within the art. In a specific embodiment, selection of antibodies that are specific to a particular domain of an GPCRX protein is facilitated by generation of hybridomas that bind to the fragment of an GPCRX protein possessing such a domain. Thus, antibodies that are specific for a desired domain within an GPCRX protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

Anti-GPCRX antibodies may be used in methods known within the art relating to the localization and/or quantitation of an GPCRX protein (*e.g.*, for use in measuring levels of the GPCRX protein within appropriate physiological samples, for use in diagnostic methods, for use in imaging the protein, and the like). In a given embodiment, antibodies for GPCRX proteins, or derivatives, fragments, analogs or homologs thereof, that contain the antibody derived binding domain, are utilized as pharmacologically-active compounds (hereinafter "Therapeutics").

An anti-GPCRX antibody (*e.g.*, monoclonal antibody) can be used to isolate an GPCRX polypeptide by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-GPCRX antibody can facilitate the purification of natural GPCRX polypeptide from cells and of recombinantly-produced GPCRX polypeptide expressed in host cells. Moreover, an anti-GPCRX antibody can be used to detect GPCRX protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the GPCRX protein. Anti-GPCRX antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

GPCRX Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding an GPCRX protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of

transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably-linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell).

The term "regulatory sequence" is intended to includes promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including

fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, GPCRX proteins, mutant forms of GPCRX proteins, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of GPCRX proteins in prokaryotic or eukaryotic cells. For example, GPCRX proteins can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (*i*) to increase expression of recombinant protein; (*ii*) to increase the solubility of the recombinant protein; and (*iii*) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein. *See, e.g.*, Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (*see, e.g.*, Wada, *et al.*, 1992. *Nucl.*

Acids Res. 20: 2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the GPCRX expression vector is a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerivisae* include pYepSec1 (Baldari, *et al.*, 1987. *EMBO J.* 6: 229-234), pMFa (Kurjan and Herskowitz, 1982. *Cell* 30: 933-943), pJRY88 (Schultz *et al.*, 1987. *Gene* 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.).

Alternatively, GPCRX can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, SF9 cells) include the pAc series (Smith, *et al.*, 1983. *Mol. Cell. Biol.* 3: 2156-2165) and the pVL series (Lucklow and Summers, 1989. *Virology* 170: 31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987. *Nature* 329: 840) and pMT2PC (Kaufman, *et al.*, 1987. *EMBO J.* 6: 187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, and simian virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see, *e.g.*, Chapters 16 and 17 of Sambrook, *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, *et al.*, 1987. *Genes Dev.* 1: 268-277), lymphoid-specific promoters (Calame and Eaton, 1988. *Adv. Immunol.* 43: 235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989. *EMBO J.* 8: 729-733) and immunoglobulins (Banerji, *et al.*, 1983. *Cell* 33: 729-740; Queen and Baltimore, 1983. *Cell* 33: 741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), pancreas-specific promoters (Edlund, *et al.*, 1985. *Science* 230: 912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, *e.g.*, the murine hox promoters (Kessel and Gruss,

1990. *Science* 249: 374-379) and the α-fetoprotein promoter (Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively-linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to GPCRX mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes *see, e.g.,* Weintraub, *et al.*, "Antisense RNA as a molecular tool for genetic analysis," *Reviews-Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, GPCRX protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory,

Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding GPCRX or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) GPCRX protein. Accordingly, the invention further provides methods for producing GPCRX protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding GPCRX protein has been introduced) in a suitable medium such that GPCRX protein is produced. In another embodiment, the method further comprises isolating GPCRX protein from the medium or the host cell.

Transgenic GPCRX Animals

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which GPCRX protein-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous GPCRX sequences have been introduced into their genome or homologous recombinant animals in which endogenous GPCRX sequences have been altered. Such animals are useful for studying the function and/or activity of GPCRX protein and for identifying and/or evaluating modulators of GPCRX protein activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and that remains in the genome of the mature animal, thereby directing the expression

of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous GPCRX gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing GPCRX-encoding nucleic acid into the male pronuclei of a fertilized oocyte (*e.g.*, by microinjection, retroviral infection) and allowing the oocyte to develop in a pseudopregnant female foster animal. The human GPCRX cDNA sequences of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84 can be introduced as a transgene into the genome of a non-human animal. Alternatively, a non-human homologue of the human GPCRX gene, such as a mouse GPCRX gene, can be isolated based on hybridization to the human GPCRX cDNA (described further *supra*) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably-linked to the GPCRX transgene to direct expression of GPCRX protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866; 4,870,009; and 4,873,191; and Hogan, 1986. In: MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the GPCRX transgene in its genome and/or expression of GPCRX mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene-encoding GPCRX protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of an GPCRX gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the GPCRX gene. The GPCRX gene can be a human gene (*e.g.*, the cDNA of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84), but more preferably, is a non-human homologue of a human GPCRX gene. For example, a mouse homologue of human GPCRX gene of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84 can be used to construct a homologous recombination vector suitable for altering an endogenous GPCRX gene in the mouse genome. In

one embodiment, the vector is designed such that, upon homologous recombination, the endogenous GPCRX gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous GPCRX gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous GPCRX protein). In the homologous recombination vector, the altered portion of the GPCRX gene is flanked at its 5'- and 3'-termini by additional nucleic acid of the GPCRX gene to allow for homologous recombination to occur between the exogenous GPCRX gene carried by the vector and an endogenous GPCRX gene in an embryonic stem cell. The additional flanking GPCRX nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5'- and 3'-termini) are included in the vector. *See, e.g.*, Thomas, *et al.*, 1987. *Cell* 51: 503 for a description of homologous recombination vectors. The vector is ten introduced into an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced GPCRX gene has homologously-recombined with the endogenous GPCRX gene are selected. *See, e.g.*, Li, *et al.*, 1992. *Cell* 69: 915.

The selected cells are then injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras. *See, e.g.*, Bradley, 1987. In: TERATOCARCINOMAS AND EMBRYONIC STEM CELLS: A PRACTICAL APPROACH, Robertson, ed. IRL, Oxford, pp. 113-152. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously-recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously-recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, 1991. *Curr. Opin. Biotechnol.* 2: 823-829; PCT International Publication Nos.: WO 90/11354; WO 91/01140; WO 92/0968; and WO 93/04169.

In another embodiment, transgenic non-humans animals can be produced that contain selected systems that allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP recombinase system, *See, e.g.*, Lakso, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae*. *See,* O'Gorman, *et al.*, 1991. *Science* 251:1351-1355. If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding

both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, *e.g.*, by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, *et al.*, 1997. *Nature* 385: 810-813. In brief, a cell (*e.g.*, a somatic cell) from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_0 phase. The quiescent cell can then be fused, *e.g.*, through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell (*e.g.*, the somatic cell) is isolated.

Pharmaceutical Compositions

The GPCRX nucleic acid molecules, GPCRX proteins, and anti-GPCRX antibodies (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and homologs thereof, can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of Remington's Pharmaceutical Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and non-aqueous vehicles such as fixed oils may also be used. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (*i.e.*, topical),

transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid (EDTA); buffers such as acetates, citrates or phosphates, and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, an GPCRX protein or anti-GPCRX antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable

solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will

be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (*see, e.g.,* U.S. Patent No. 5,328,470) or by stereotactic injection (*see, e.g.*, Chen, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells that produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

Screening and Detection Methods

The isolated nucleic acid molecules of the invention can be used to express GPCRX protein (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect GPCRX mRNA (*e.g.*, in a biological sample) or a genetic lesion in an GPCRX gene, and to modulate GPCRX activity, as described further, below. In addition, the GPCRX proteins can be used to screen drugs or compounds that modulate the GPCRX protein activity or expression as well as to treat disorders characterized by insufficient or excessive production of GPCRX protein or production of GPCRX protein forms that have decreased or aberrant activity compared

to GPCRX wild-type protein (*e.g.*; diabetes (regulates insulin release); obesity (binds and transport lipids); metabolic disturbances associated with obesity, the metabolic syndrome X as well as anorexia and wasting disorders associated with chronic diseases and various cancers, and infectious disease(possesses anti-microbial activity) and the various dyslipidemias. In addition, the anti-GPCRX antibodies of the invention can be used to detect and isolate GPCRX proteins and modulate GPCRX activity. In yet a further aspect, the invention can be used in methods to influence appetite, absorption of nutrients and the disposition of metabolic substrates in both a positive and negative fashion.

The invention further pertains to novel agents identified by the screening assays described herein and uses thereof for treatments as described, *supra*.

Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) that bind to GPCRX proteins or have a stimulatory or inhibitory effect on, *e.g.*, GPCRX protein expression or GPCRX protein activity. The invention also includes compounds identified in the screening assays described herein.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form of an GPCRX protein or polypeptide or biologically-active portion thereof. The test compounds of the invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds. *See, e.g.,* Lam, 1997. *Anticancer Drug Design* 12: 145.

A "small molecule" as used herein, is meant to refer to a composition that has a molecular weight of less than about 5 kD and most preferably less than about 4 kD. Small molecules can be, *e.g.*, nucleic acids, peptides, polypeptides, peptidomimetics, carbohydrates, lipids or other organic or inorganic molecules. Libraries of chemical and/or biological mixtures, such as fungal, bacterial, or algal extracts, are known in the art and can be screened with any of the assays of the invention.

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt, *et al.*, 1993. *Proc. Natl. Acad. Sci. U.S.A.* 90: 6909; Erb, *et al.*, 1994. *Proc. Natl. Acad. Sci. U.S.A.* 91: 11422; Zuckermann, *et al.*, 1994. *J. Med. Chem.* 37: 2678; Cho, *et al.*, 1993. *Science* 261: 1303; Carrell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2059; Carell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2061; and Gallop, *et al.*, 1994. *J. Med. Chem.* 37: 1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten, 1992. *Biotechniques* 13: 412-421), or on beads (Lam, 1991. *Nature* 354: 82-84), on chips (Fodor, 1993. *Nature* 364: 555-556), bacteria (Ladner, U.S. Patent No. 5,223,409), spores (Ladner, U.S. Patent 5,233,409), plasmids (Cull, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 1865-1869) or on phage (Scott and Smith, 1990. *Science* 249: 386-390; Devlin, 1990. *Science* 249: 404-406; Cwirla, *et al.*, 1990. *Proc. Natl. Acad. Sci. U.S.A.* 87: 6378-6382; Felici, 1991. *J. Mol. Biol.* 222: 301-310; Ladner, U.S. Patent No. 5,233,409.).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of GPCRX protein, or a biologically-active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to an GPCRX protein determined. The cell, for example, can of mammalian origin or a yeast cell. Determining the ability of the test compound to bind to the GPCRX protein can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the GPCRX protein or biologically-active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, test compounds can be enzymatically-labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In one embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of GPCRX protein, or a biologically-active portion thereof, on the cell surface with a known compound which binds GPCRX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with an GPCRX protein, wherein determining the ability of the test compound to interact with an GPCRX protein comprises determining the ability of the test compound to preferentially bind to GPCRX protein or a biologically-active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of GPCRX protein, or a biologically-active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the GPCRX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of GPCRX or a biologically-active portion thereof can be accomplished, for example, by determining the ability of the GPCRX protein to bind to or interact with an GPCRX target molecule. As used herein, a "target molecule" is a molecule with which an GPCRX protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses an GPCRX interacting protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. An GPCRX target molecule can be a non-GPCRX molecule or an GPCRX protein or polypeptide of the invention. In one embodiment, an GPCRX target molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (*e.g.* a signal generated by binding of a compound to a membrane-bound GPCRX molecule) through the cell membrane and into the cell. The target, for example, can be a second intercellular protein that has catalytic activity or a protein that facilitates the association of downstream signaling molecules with GPCRX.

Determining the ability of the GPCRX protein to bind to or interact with an GPCRX target molecule can be accomplished by one of the methods described above for determining direct binding. In one embodiment, determining the ability of the GPCRX protein to bind to or interact with an GPCRX target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*i.e.* intracellular Ca^{2+}, diacylglycerol, IP_3, etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene (comprising an GPCRX-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the invention is a cell-free assay comprising contacting an GPCRX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to bind to the GPCRX protein or biologically-active portion thereof. Binding of the test compound to the GPCRX protein can be determined either directly or indirectly as described above. In one such embodiment, the assay comprises contacting the GPCRX protein or biologically-active portion thereof with a known compound

which binds GPCRX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with an GPCRX protein, wherein determining the ability of the test compound to interact with an GPCRX protein comprises determining the ability of the test compound to preferentially bind to GPCRX or biologically-active portion thereof as compared to the known compound.

In still another embodiment, an assay is a cell-free assay comprising contacting GPCRX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to modulate (*e.g.* stimulate or inhibit) the activity of the GPCRX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of GPCRX can be accomplished, for example, by determining the ability of the GPCRX protein to bind to an GPCRX target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of GPCRX protein can be accomplished by determining the ability of the GPCRX protein further modulate an GPCRX target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as described, *supra*.

In yet another embodiment, the cell-free assay comprises contacting the GPCRX protein or biologically-active portion thereof with a known compound which binds GPCRX protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with an GPCRX protein, wherein determining the ability of the test compound to interact with an GPCRX protein comprises determining the ability of the GPCRX protein to preferentially bind to or modulate the activity of an GPCRX target molecule.

The cell-free assays of the invention are amenable to use of both the soluble form or the membrane-bound form of GPCRX protein. In the case of cell-free assays comprising the membrane-bound form of GPCRX protein, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of GPCRX protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)$_n$, N-dodecyl--N,N-dimethyl-3-ammonio-1-propane sulfonate, 3-(3-cholamidopropyl) dimethylamminiol-1-propane sulfonate (CHAPS), or 3-(3-cholamidopropyl)dimethylamminiol-2-hydroxy-1-propane sulfonate (CHAPSO).

In more than one embodiment of the above assay methods of the invention, it may be desirable to immobilize either GPCRX protein or its target molecule to facilitate separation of

complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to GPCRX protein, or interaction of GPCRX protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For example, GST-GPCRX fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, that are then combined with the test compound or the test compound and either the non-adsorbed target protein or GPCRX protein, and the mixture is incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described, *supra*. Alternatively, the complexes can be dissociated from the matrix, and the level of GPCRX protein binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the GPCRX protein or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated GPCRX protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well-known within the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with GPCRX protein or target molecules, but which do not interfere with binding of the GPCRX protein to its target molecule, can be derivatized to the wells of the plate, and unbound target or GPCRX protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the GPCRX protein or target molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity associated with the GPCRX protein or target molecule.

In another embodiment, modulators of GPCRX protein expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of GPCRX mRNA or protein in the cell is determined. The level of expression of GPCRX mRNA or protein in the presence of the candidate compound is compared to the level of expression of GPCRX mRNA or protein in the absence of the candidate compound. The candidate compound can then

be identified as a modulator of GPCRX mRNA or protein expression based upon this comparison. For example, when expression of GPCRX mRNA or protein is greater (*i.e.*, statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of GPCRX mRNA or protein expression. Alternatively, when expression of GPCRX mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of GPCRX mRNA or protein expression. The level of GPCRX mRNA or protein expression in the cells can be determined by methods described herein for detecting GPCRX mRNA or protein.

In yet another aspect of the invention, the GPCRX proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos, *et al.*, 1993. *Cell* 72: 223-232; Madura, *et al.*, 1993. *J. Biol. Chem.* 268: 12046-12054; Bartel, *et al.*, 1993. *Biotechniques* 14: 920-924; Iwabuchi, *et al.*, 1993. *Oncogene* 8: 1693-1696; and Brent WO 94/10300), to identify other proteins that bind to or interact with GPCRX ("GPCRX-binding proteins" or "GPCRX-bp") and modulate GPCRX activity. Such GPCRX-binding proteins are also likely to be involved in the propagation of signals by the GPCRX proteins as, for example, upstream or downstream elements of the GPCRX pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for GPCRX is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming an GPCRX-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein which interacts with GPCRX.

The invention further pertains to novel agents identified by the aforementioned screening assays and uses thereof for treatments as described herein.

Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. By way of example, and not of limitation, these sequences can be used to: (*i*) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (*ii*) identify an individual from a minute biological sample (tissue typing); and (*iii*) aid in forensic identification of a biological sample. Some of these applications are described in the subsections, below.

Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the GPCRX sequences, SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84, or fragments or derivatives thereof, can be used to map the location of the GPCRX genes, respectively, on a chromosome. The mapping of the GPCRX sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, GPCRX genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the GPCRX sequences. Computer analysis of the GPCRX, sequences can be used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the GPCRX sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (*e.g.*, human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but in which human cells can, the one human chromosome that contains the gene encoding the needed enzyme will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. *See, e.g.,* D'Eustachio, *et al.*, 1983. *Science* 220: 919-924. Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the GPCRX sequences to design oligonucleotide primers, sub-localization can be achieved with panels of fragments from specific chromosomes.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical like colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases, will suffice to get good results at a reasonable amount of time. For a review of this technique, *see,* Verma, *et al.*, HUMAN CHROMOSOMES: A MANUAL OF BASIC TECHNIQUES (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, *e.g.*, in McKusick, MENDELIAN INHERITANCE IN MAN, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, *e.g.*, Egeland, *et al.*, 1987. *Nature*, 325: 783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the GPCRX gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the

chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Tissue Typing

The GPCRX sequences of the invention can also be used to identify individuals from minute biological samples. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. The sequences of the invention are useful as additional DNA markers for RFLP ("restriction fragment length polymorphisms," described in U.S. Patent No. 5,272,057).

Furthermore, the sequences of the invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the GPCRX sequences described herein can be used to prepare two PCR primers from the 5'- and 3'-termini of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the invention can be used to obtain such identification sequences from individuals and from tissue. The GPCRX sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Much of the allelic variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms (RFLPs).

Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

Predictive Medicine

The invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the invention relates to diagnostic assays for determining GPCRX protein and/or nucleic acid expression as well as GPCRX activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant GPCRX expression or activity. The disorders include metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with GPCRX protein, nucleic acid expression or activity. For example, mutations in an GPCRX gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with GPCRX protein, nucleic acid expression, or biological activity.

Another aspect of the invention provides methods for determining GPCRX protein, nucleic acid expression or activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of GPCRX in clinical trials.

These and other agents are described in further detail in the following sections.

Diagnostic Assays

An exemplary method for detecting the presence or absence of GPCRX in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting GPCRX protein or nucleic acid (*e.g.*,

mRNA, genomic DNA) that encodes GPCRX protein such that the presence of GPCRX is detected in the biological sample. An agent for detecting GPCRX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to GPCRX mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length GPCRX nucleic acid, such as the nucleic acid of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 28, 30, 32, 34, 36, 38, and 84, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to GPCRX mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

An agent for detecting GPCRX protein is an antibody capable of binding to GPCRX protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect GPCRX mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of GPCRX mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of GPCRX protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of GPCRX genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of GPCRX protein include introducing into a subject a labeled anti-GPCRX antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting GPCRX protein, mRNA, or genomic DNA, such that the presence of GPCRX protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of GPCRX protein, mRNA or genomic DNA in the control sample with the presence of GPCRX protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of GPCRX in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting GPCRX protein or mRNA in a biological sample; means for determining the amount of GPCRX in the sample; and means for comparing the amount of GPCRX in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect GPCRX protein or nucleic acid.

Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant GPCRX expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with GPCRX protein, nucleic acid expression or activity. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disease or disorder. Thus, the invention provides a method for identifying a disease or disorder associated with aberrant GPCRX expression or activity in which a test sample is obtained from a subject and GPCRX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) is detected, wherein the presence of GPCRX protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant GPCRX expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant GPCRX expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder. Thus, the invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant GPCRX expression or activity in which a test

sample is obtained and GPCRX protein or nucleic acid is detected (*e.g.*, wherein the presence of GPCRX protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant GPCRX expression or activity).

The methods of the invention can also be used to detect genetic lesions in an GPCRX gene, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized by aberrant cell proliferation and/or differentiation. In various embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of an alteration affecting the integrity of a gene encoding an GPCRX-protein, or the misexpression of the GPCRX gene. For example, such genetic lesions can be detected by ascertaining the existence of at least one of: (*i*) a deletion of one or more nucleotides from an GPCRX gene; (*ii*) an addition of one or more nucleotides to an GPCRX gene; (*iii*) a substitution of one or more nucleotides of an GPCRX gene, (*iv*) a chromosomal rearrangement of an GPCRX gene; (*v*) an alteration in the level of a messenger RNA transcript of an GPCRX gene, (*vi*) aberrant modification of an GPCRX gene, such as of the methylation pattern of the genomic DNA, (*vii*) the presence of a non-wild-type splicing pattern of a messenger RNA transcript of an GPCRX gene, (*viii*) a non-wild-type level of an GPCRX protein, (*ix*) allelic loss of an GPCRX gene, and (*x*) inappropriate post-translational modification of an GPCRX protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in an GPCRX gene. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (*see, e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (*see, e.g.*, Landegran, *et al.*, 1988. *Science* 241: 1077-1080; and Nakazawa, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 360-364), the latter of which can be particularly useful for detecting point mutations in the GPCRX-gene (*see,* Abravaya, *et al.*, 1995. *Nucl. Acids Res.* 23: 675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers that specifically hybridize to an GPCRX gene under conditions such that hybridization and amplification of the GPCRX gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR

may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (*see,* Guatelli, *et al.*, 1990. *Proc. Natl. Acad. Sci. USA* 87: 1874-1878), transcriptional amplification system (*see,* Kwoh, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 1173-1177); Qβ Replicase (*see,* Lizardi, *et al*, 1988. *BioTechnology* 6: 1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in an GPCRX gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (*see, e.g.,* U.S. Patent No. 5,493,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in GPCRX can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high-density arrays containing hundreds or thousands of oligonucleotides probes. *See, e.g.,* Cronin, *et al.*, 1996. *Human Mutation* 7: 244-255; Kozal, *et al.*, 1996. *Nat. Med.* 2: 753-759. For example, genetic mutations in GPCRX can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, *et al.*, *supra.* Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the GPCRX gene and detect mutations by comparing the sequence of the sample GPCRX with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert, 1977. *Proc. Natl. Acad. Sci. USA* 74: 560 or Sanger, 1977. *Proc. Natl. Acad. Sci. USA* 74: 5463. It is

also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (*see, e.g.,* Naeve, *et al.*, 1995. *Biotechniques* 19: 448), including sequencing by mass spectrometry (see, *e.g.*. PCT International Publication No. WO 94/16101; Cohen, *et al.*, 1996. *Adv. Chromatography* 36: 127-162; and Griffin, *et al.*, 1993. *Appl. Biochem. Biotechnol.* 38: 147-159).

Other methods for detecting mutations in the GPCRX gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes. *See, e.g.,* Myers, *et al.*, 1985. *Science* 230: 1242. In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type GPCRX sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S_1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. *See, e.g.,* Cotton, *et al.*, 1988. *Proc. Natl. Acad. Sci. USA* 85: 4397; Saleeba, *et al.*, 1992. *Methods Enzymol.* 217: 286-295. In an embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in GPCRX cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches. *See, e.g.,* Hsu, *et al.*, 1994. *Carcinogenesis* 15: 1657-1662. According to an exemplary embodiment, a probe based on an GPCRX sequence, *e.g.*, a wild-type GPCRX sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. *See, e.g.,* U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in GPCRX genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type

nucleic acids. *See, e.g.,* Orita, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA*: 86: 2766; Cotton, 1993. *Mutat. Res.* 285: 125-144; Hayashi, 1992. *Genet. Anal. Tech. Appl.* 9: 73-79. Single-stranded DNA fragments of sample and control GPCRX nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. *See, e.g.,* Keen, *et al.*, 1991. *Trends Genet.* 7: 5.

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE). *See. e.g.,* Myers, *et al.*, 1985. *Nature* 313: 495. When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA. *See, e.g.,* Rosenbaum and Reissner, 1987. *Biophys. Chem.* 265: 12753.

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions that permit hybridization only if a perfect match is found. *See, e.g.,* Saiki, *et al.*, 1986. *Nature* 324: 163; Saiki, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 6230. Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization; *see, e.g.,* Gibbs, *et al.*, 1989. *Nucl. Acids Res.* 17: 2437-2448) or at the extreme 3'-terminus of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (*see, e.g.,* Prossner, 1993. *Tibtech.* 11: 238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection. *See, e.g.,* Gasparini, *et al.*, 1992. *Mol.*

Cell Probes 6: 1. It is anticipated that in certain embodiments amplification may also be performed using *Taq* ligase for amplification. *See, e.g.,* Barany, 1991. *Proc. Natl. Acad. Sci. USA* 88: 189. In such cases, ligation will occur only if there is a perfect match at the 3'-terminus of the 5' sequence, making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving an GPCRX gene.

Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which GPCRX is expressed may be utilized in the prognostic assays described herein. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

Pharmacogenomics

Agents, or modulators that have a stimulatory or inhibitory effect on GPCRX activity (*e.g.*, GPCRX gene expression), as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders (The disorders include metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers.) In conjunction with such treatment, the pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (*e.g.*, drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of GPCRX protein, expression of GPCRX nucleic acid, or mutation content of GPCRX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See *e.g.*, Eichelbaum, 1996. *Clin. Exp. Pharmacol. Physiol.*, 23: 983-985; Linder, 1997. *Clin. Chem.*, 43: 254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. At the other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of GPCRX protein, expression of GPCRX nucleic acid, or mutation content of GPCRX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with an GPCRX

modulator, such as a modulator identified by one of the exemplary screening assays described herein.

Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of GPCRX (*e.g.*, the ability to modulate aberrant cell proliferation and/or differentiation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase GPCRX gene expression, protein levels, or upregulate GPCRX activity, can be monitored in clinical trails of subjects exhibiting decreased GPCRX gene expression, protein levels, or downregulated GPCRX activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease GPCRX gene expression, protein levels, or downregulate GPCRX activity, can be monitored in clinical trails of subjects exhibiting increased GPCRX gene expression, protein levels, or upregulated GPCRX activity. In such clinical trials, the expression or activity of GPCRX and, preferably, other genes that have been implicated in, for example, a cellular proliferation or immune disorder can be used as a "read out" or markers of the immune responsiveness of a particular cell.

By way of example, and not of limitation, genes, including GPCRX, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) that modulates GPCRX activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of GPCRX and other genes implicated in the disorder. The levels of gene expression (*i.e.*, a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of GPCRX or other genes. In this manner, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In one embodiment, the invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, protein, peptide, peptidomimetic, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (*i*) obtaining a pre-administration sample from a subject prior to administration of the agent; (*ii*) detecting the level of expression of an GPCRX

protein, mRNA, or genomic DNA in the preadministration sample; (*iii*) obtaining one or more post-administration samples from the subject; (*iv*) detecting the level of expression or activity of the GPCRX protein, mRNA, or genomic DNA in the post-administration samples; (*v*) comparing the level of expression or activity of the GPCRX protein, mRNA, or genomic DNA in the pre-administration sample with the GPCRX protein, mRNA, or genomic DNA in the post administration sample or samples; and (*vi*) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of GPCRX to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of GPCRX to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

Methods of Treatment

The invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant GPCRX expression or activity. The disorders include cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, transplantation, adrenoleukodystrophy, congenital adrenal hyperplasia, prostate cancer, neoplasm; adenocarcinoma, lymphoma, uterus cancer, fertility, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, graft versus host disease, AIDS, bronchial asthma, Crohn's disease; multiple sclerosis, treatment of Albright Hereditary Ostoeodystrophy, and other diseases, disorders and conditions of the like.

These methods of treatment will be discussed more fully, below.

Disease and Disorders

Diseases and disorders that are characterized by increased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that antagonize (*i.e.*, reduce or inhibit) activity. Therapeutics that antagonize activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to: (*i*) an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; (*ii*) antibodies to an aforementioned peptide; (*iii*) nucleic acids encoding an aforementioned peptide; (*iv*) administration of antisense nucleic acid and nucleic

acids that are "dysfunctional" (*i.e.*, due to a heterologous insertion within the coding sequences of coding sequences to an aforementioned peptide) that are utilized to "knockout" endoggenous function of an aforementioned peptide by homologous recombination (*see, e.g.*, Capecchi, 1989. *Science* 244: 1288-1292); or (*v*) modulators (*i.e.*, inhibitors, agonists and antagonists, including additional peptide mimetic of the invention or antibodies specific to a peptide of the invention) that alter the interaction between an aforementioned peptide and its binding partner.

Diseases and disorders that are characterized by decreased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that increase (*i.e.*, are agonists to) activity. Therapeutics that upregulate activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to, an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

Increased or decreased levels can be readily detected by quantifying peptide and/or RNA, by obtaining a patient tissue sample (*e.g.*, from biopsy tissue) and assaying it *in vitro* for RNA or peptide levels, structure and/or activity of the expressed peptides (or mRNAs of an aforementioned peptide). Methods that are well-known within the art include, but are not limited to, immunoassays (*e.g.*, by Western blot analysis, immunoprecipitation followed by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, immunocytochemistry, etc.) and/or hybridization assays to detect expression of mRNAs (*e.g.*, Northern assays, dot blots, *in situ* hybridization, and the like).

Prophylactic Methods

In one aspect, the invention provides a method for preventing, in a subject, a disease or condition associated with an aberrant GPCRX expression or activity, by administering to the subject an agent that modulates GPCRX expression or at least one GPCRX activity. Subjects at risk for a disease that is caused or contributed to by aberrant GPCRX expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the GPCRX aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending upon the type of GPCRX aberrancy, for example, an GPCRX agonist or GPCRX antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein. The prophylactic methods of the invention are further discussed in the following subsections.

Therapeutic Methods

Another aspect of the invention pertains to methods of modulating GPCRX expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of GPCRX protein activity associated with the cell. An agent that modulates GPCRX protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of an GPCRX protein, a peptide, an GPCRX peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more GPCRX protein activity. Examples of such stimulatory agents include active GPCRX protein and a nucleic acid molecule encoding GPCRX that has been introduced into the cell. In another embodiment, the agent inhibits one or more GPCRX protein activity. Examples of such inhibitory agents include antisense GPCRX nucleic acid molecules and anti-GPCRX antibodies. These modulatory methods can be performed *in vitro* (*e.g.*, by culturing the cell with the agent) or, alternatively, *in vivo* (*e.g.*, by administering the agent to a subject). As such, the invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of an GPCRX protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (*e.g.*, an agent identified by a screening assay described herein), or combination of agents that modulates (*e.g.*, up-regulates or down-regulates) GPCRX expression or activity. In another embodiment, the method involves administering an GPCRX protein or nucleic acid molecule as therapy to compensate for reduced or aberrant GPCRX expression or activity.

Stimulation of GPCRX activity is desirable in situations in which GPCRX is abnormally downregulated and/or in which increased GPCRX activity is likely to have a beneficial effect. One example of such a situation is where a subject has a disorder characterized by aberrant cell proliferation and/or differentiation (*e.g.*, cancer or immune associated disorders). Another example of such a situation is where the subject has a gestational disease (*e.g.*, preclampsia).

Determination of the Biological Effect of the Therapeutic

In various embodiments of the invention, suitable *in vitro* or *in vivo* assays are performed to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected tissue.

In various specific embodiments, *in vitro* assays may be performed with representative cells of the type(s) involved in the patient's disorder, to determine if a given Therapeutic exerts the desired effect upon the cell type(s). Compounds for use in therapy may be tested in suitable

animal model systems including, but not limited to rats, mice, chicken, cows, monkeys, rabbits, and the like, prior to testing in human subjects. Similarly, for *in vivo* testing, any of the animal model system known in the art may be used prior to administration to human subjects.

Prophylactic and Therapeutic Uses of the Compositions of the Invention

The GPCRX nucleic acids and proteins of the invention are useful in potential prophylactic and therapeutic applications implicated in a variety of disorders including, but not limited to: metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers.

As an example, a cDNA encoding the GPCRX protein of the invention may be useful in gene therapy, and the protein may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the invention will have efficacy for treatment of patients suffering from: metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, hematopoietic disorders, and the various dyslipidemias.

Both the novel nucleic acid encoding the GPCRX protein, and the GPCRX protein of the invention, or fragments thereof, may also be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. A further use could be as an anti-bacterial molecule (*i.e.,* some peptides have been found to possess anti-bacterial properties). These materials are further useful in the generation of antibodies which immunospecifically-bind to the novel substances of the invention for use in therapeutic or diagnostic methods.

The invention will be further illustrated in the following non-limiting examples.

EXAMPLES

Example 1 - Quantitative Expression Analysis of Clones in Various Cells and Tissues

The quantitative expression of various clones was assessed using microtiter plates containing RNA samples from a variety of normal and pathology-derived cells, cell lines and tissues using real time quantitative PCR (RTQ PCR; TAQMAN®). RTQ PCR was performed on a Perkin-Elmer Biosystems ABI PRISM® 7700 Sequence Detection System. Various collections of samples are assembled on the plates, and referred to as Panel 1 (containing cells

and cell lines from normal and cancer sources), Panel 2 (containing samples derived from tissues, in particular from surgical samples, from normal and cancer sources), Panel 3 (containing samples derived from a wide variety of cancer sources), Panel 4 (containing cells and cell lines from normal cells and cells related to inflammatory conditions) and Panel CNSD.01 (containing samples from normal and diseased brains).

First, the RNA samples were normalized to constitutively expressed genes such as b-actin and GAPDH. RNA (~50 ng total or ~1 ng polyA+) was converted to cDNA using the TAQMAN® Reverse Transcription Reagents Kit (PE Biosystems, Foster City, CA; Catalog No. N808-0234) and random hexamers according to the manufacturer's protocol. Reactions were performed in 20 ul and incubated for 30 min. at 480C. cDNA (5 ul) was then transferred to a separate plate for the TAQMAN® reaction using b-actin and GAPDH TAQMAN® Assay Reagents (PE Biosystems; Catalog Nos. 4310881E and 4310884E, respectively) and TAQMAN® universal PCR Master Mix (PE Biosystems; Catalog No. 4304447) according to the manufacturer's protocol. Reactions were performed in 25 ul using the following parameters: 2 min. at 500C; 10 min. at 950C; 15 sec. at 950C/1 min. at 600C (40 cycles). Results were recorded as CT values (cycle at which a given sample crosses a threshold level of fluorescence) using a log scale, with the difference in RNA concentration between a given sample and the sample with the lowest CT value being represented as 2 to the power of delta CT. The percent relative expression is then obtained by taking the reciprocal of this RNA difference and multiplying by 100. The average CT values obtained for ß-actin and GAPDH were used to normalize RNA samples. The RNA sample generating the highest CT value required no further diluting, while all other samples were diluted relative to this sample according to their b-actin /GAPDH average CT values.

Normalized RNA (5 ul) was converted to cDNA and analyzed via TAQMAN® using One Step RT-PCR Master Mix Reagents (PE Biosystems; Catalog No. 4309169) and gene-specific primers according to the manufacturer's instructions. Probes and primers were designed for each assay according to Perkin Elmer Biosystem's Primer Express Software package (version I for Apple Computer's Macintosh Power PC) or a similar algorithm using the target sequence as input. Default settings were used for reaction conditions and the following parameters were set before selecting primers: primer concentration = 250 nM, primer melting temperature (Tm) range = 58°-60° C, primer optimal Tm = 59° C, maximum primer difference = 2° C, probe does not have 5' G, probe Tm must be 10° C greater than primer Tm, amplicon size 75 bp to 100 bp. The probes and primers selected (see below) were synthesized by Synthegen (Houston, TX, USA). Probes were double purified by HPLC to remove uncoupled dye and evaluated by mass

spectroscopy to verify coupling of reporter and quencher dyes to the 5' and 3' ends of the probe, respectively. Their final concentrations were: forward and reverse primers, 900 nM each, and probe, 200nM.

PCR conditions: Normalized RNA from each tissue and each cell line was spotted in each well of a 96 well PCR plate (Perkin Elmer Biosystems). PCR cocktails including two probes (a probe specific for the target clone and another gene-specific probe multiplexed with the target probe) were set up using 1X TaqManÔ PCR Master Mix for the PE Biosystems 7700, with 5 mM MgCl2, dNTPs (dA, G, C, U at 1:1:1:2 ratios), 0.25 U/ml AmpliTaq GoldÔ (PE Biosystems), and 0.4 U/ml RNase inhibitor, and 0.25 U/ml reverse transcriptase. Reverse transcription was performed at 48° C for 30 minutes followed by amplification/PCR cycles as follows: 95° C 10 min, then 40 cycles of 95° C for 15 seconds, 60° C for 1 minute.

In the results for Panel 1, the following abbreviations are used:

ca. = carcinoma,

* = established from metastasis,

met = metastasis,

s cell var= small cell variant,

non-s = non-sm =non-small,

squam = squamous,

pl. eff = pl effusion = pleural effusion,

glio = glioma,

astro = astrocytoma, and

neuro = neuroblastoma.

Panel 2

The plates for Panel 2 generally include 2 control wells and 94 test samples composed of RNA or cDNA isolated from human tissue procured by surgeons working in close cooperation with the National Cancer Institute's Cooperative Human Tissue Network (CHTN) or the National Disease Research Initiative (NDRI). The tissues are derived from human malignancies and in cases where indicated many malignant tissues have "matched margins" obtained from noncancerous tissue just adjacent to the tumor. These are termed normal adjacent tissues and are denoted "NAT" in the results below. The tumor tissue and the "matched margins" are evaluated by two independent pathologists (the surgical pathologists and again by a pathologists at NDRI or CHTN). This analysis provides a gross histopathological assessment of tumor differentiation

grade. Moreover, most samples include the original surgical pathology report that provides information regarding the clinical stage of the patient. These matched margins are taken from the tissue surrounding (i.e. immediately proximal) to the zone of surgery (designated "NAT", for normal adjacent tissue, in Table RR). In addition, RNA and cDNA samples were obtained from various human tissues derived from autopsies performed on elderly people or sudden death victims (accidents, etc.). These tissues were ascertained to be free of disease and were purchased from various commercial sources such as Clontech (Palo Alto, CA), Research Genetics, and Invitrogen.

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

Panel 3D

The plates of Panel 3D are comprised of 94 cDNA samples and two control samples. Specifically, 92 of these samples are derived from cultured human cancer cell lines, 2 samples of human primary cerebellar tissue and 2 controls. The human cell lines are generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups: Squamous cell carcinoma of the tongue, breast cancer, prostate cancer, melanoma, epidermoid carcinoma, sarcomas, bladder carcinomas, pancreatic cancers, kidney cancers, leukemias/lymphomas, ovarian/uterine/cervical, gastric, colon, lung and CNS cancer cell lines. In addition, there are two independent samples of cerebellum. These cells are all cultured under standard recommended conditions and RNA extracted using the standard procedures. The cell lines in panel 3D and 1.3D are of the most common cell lines used in the scientific literature.

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

Panel 4

Panel 4 includes samples on a 96 well plate (2 control wells, 94 test samples) composed of RNA (Panel 4r) or cDNA (Panel 4d) isolated from various human cell lines or tissues related to inflammatory conditions. Total RNA from control normal tissues such as colon and lung (Stratagene ,La Jolla, CA) and thymus and kidney (Clontech) were employed. Total RNA from liver tissue from cirrhosis patients and kidney from lupus patients was obtained from BioChain (Biochain Institute, Inc., Hayward, CA). Intestinal tissue for RNA preparation from patients diagnosed as having Crohn's disease and ulcerative colitis was obtained from the National Disease Research Interchange (NDRI) (Philadelphia, PA).

Astrocytes, lung fibroblasts, dermal fibroblasts, coronary artery smooth muscle cells, small airway epithelium, bronchial epithelium, microvascular dermal endothelial cells, microvascular lung endothelial cells, human pulmonary aortic endothelial cells, human umbilical vein endothelial cells were all purchased from Clonetics (Walkersville, MD) and grown in the media supplied for these cell types by Clonetics. These primary cell types were activated with various cytokines or combinations of cytokines for 6 and/or 12-14 hours, as indicated. The following cytokines were used; IL-1 beta at approximately 1-5 ng/ml, TNF alpha at approximately 5-10 ng/ml, IFN gamma at approximately 20-50 ng/ml, IL-4 at approximately 5-10 ng/ml, IL-9 at approximately 5-10 ng/ml, IL-13 at approximately 5-10 ng/ml. Endothelial cells were sometimes starved for various times by culture in the basal media from Clonetics with 0.1% serum.

Mononuclear cells were prepared from blood of employees at CuraGen Corporation, using Ficoll. LAK cells were prepared from these cells by culture in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco/Life Technologies, Rockville, MD), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10-5 M (Gibco), and 10 mM Hepes (Gibco) and Interleukin 2 for 4-6 days. Cells were then either activated with 10-20 ng/ml PMA and 1-2 mg/ml ionomycin, IL-12 at 5-10 ng/ml, IFN gamma at 20-50 ng/ml and IL-18 at 5-10 ng/ml for 6 hours. In some cases, mononuclear cells were cultured for 4-5 days in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10-5 M (Gibco), and 10 mM Hepes (Gibco) with PHA (phytohemagglutinin) or PWM (pokeweed mitogen) at approximately 5 mg/ml. Samples were taken at 24, 48 and 72 hours for RNA preparation. MLR (mixed lymphocyte reaction) samples were obtained by taking blood from two donors, isolating the mononuclear cells using Ficoll and mixing the isolated mononuclear cells 1:1 at a final concentration of approximately 2x106 cells/ml in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1 mM

sodium pyruvate (Gibco), mercaptoethanol (5.5 x 10-5 M) (Gibco), and 10 mM Hepes (Gibco). The MLR was cultured and samples taken at various time points ranging from 1- 7 days for RNA preparation.

Monocytes were isolated from mononuclear cells using CD14 Miltenyi Beads, +ve VS selection columns and a Vario Magnet according to the manufacturer's instructions. Monocytes were differentiated into dendritic cells by culture in DMEM 5% fetal calf serum (FCS) (Hyclone, Logan, UT), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10-5 M (Gibco), and 10 mM Hepes (Gibco), 50 ng/ml GMCSF and 5 ng/ml IL-4 for 5-7 days. Macrophages were prepared by culture of monocytes for 5-7 days in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10-5 M (Gibco), 10 mM Hepes (Gibco) and 10% AB Human Serum or MCSF at approximately 50 ng/ml. Monocytes, macrophages and dendritic cells were stimulated for 6 and 12-14 hours with lipopolysaccharide (LPS) at 100 ng/ml. Dendritic cells were also stimulated with anti-CD40 monoclonal antibody (Pharmingen) at 10 mg/ml for 6 and 12-14 hours.

CD4 lymphocytes, CD8 lymphocytes and NK cells were also isolated from mononuclear cells using CD4, CD8 and CD56 Miltenyi beads, positive VS selection columns and a Vario Magnet according to the manufacturer's instructions. CD45RA and CD45RO CD4 lymphocytes were isolated by depleting mononuclear cells of CD8, CD56, CD14 and CD19 cells using CD8, CD56, CD14 and CD19 Miltenyi beads and positive selection. Then CD45RO beads were used to isolate the CD45RO CD4 lymphocytes with the remaining cells being CD45RA CD4 lymphocytes. CD45RA CD4, CD45RO CD4 and CD8 lymphocytes were placed in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10-5 M (Gibco), and 10 mM Hepes (Gibco) and plated at 106 cells/ml onto Falcon 6 well tissue culture plates that had been coated overnight with 0.5 mg/ml anti-CD28 (Pharmingen) and 3 ug/ml anti-CD3 (OKT3, ATCC) in PBS. After 6 and 24 hours, the cells were harvested for RNA preparation. To prepare chronically activated CD8 lymphocytes, we activated the isolated CD8 lymphocytes for 4 days on anti-CD28 and anti-CD3 coated plates and then harvested the cells and expanded them in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10-5 M (Gibco), and 10 mM Hepes (Gibco) and IL-2. The expanded CD8 cells were then activated again with plate bound anti-CD3 and anti-CD28 for 4 days and expanded as before. RNA was isolated 6 and 24 hours after the second activation and after 4 days of the second expansion culture. The isolated NK cells were cultured in DMEM 5% FCS (Hyclone), 100 mM non essential amino

acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10-5 M (Gibco), and 10 mM Hepes (Gibco) and IL-2 for 4-6 days before RNA was prepared.

To obtain B cells, tonsils were procured from NDRI. The tonsil was cut up with sterile dissecting scissors and then passed through a sieve. Tonsil cells were then spun down and resuspended at 106 cells/ml in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10-5 M (Gibco), and 10 mM Hepes (Gibco). To activate the cells, we used PWM at 5 mg/ml or anti-CD40 (Pharmingen) at approximately 10 mg/ml and IL-4 at 5-10 ng/ml. Cells were harvested for RNA preparation at 24,48 and 72 hours.

To prepare the primary and secondary Th1/Th2 and Tr1 cells, six-well Falcon plates were coated overnight with 10 μg/ml anti-CD28 (Pharmingen) and 2 μg/ml OKT3 (ATCC), and then washed twice with PBS. Umbilical cord blood CD4 lymphocytes (Poietic Systems, German Town, MD) were cultured at 105-106 cells/ml in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10-5 M (Gibco), 10 mM Hepes (Gibco) and IL-2 (4 ng/ml). IL-12 (5 ng/ml) and anti-IL4 (1 g/ml) were used to direct to Th1, while IL-4 (5 ng/ml) and anti-IFN gamma (1 g/ml) were used to direct to Th2 and IL-10 at 5 ng/ml was used to direct to Tr1. After 4-5 days, the activated Th1, Th2 and Tr1 lymphocytes were washed once in DMEM and expanded for 4-7 days in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10-5 M (Gibco), 10 mM Hepes (Gibco) and IL-2 (1 ng/ml). Following this, the activated Th1, Th2 and Tr1 lymphocytes were re-stimulated for 5 days with anti-CD28/OKT3 and cytokines as described above, but with the addition of anti-CD95L (1 g/ml) to prevent apoptosis. After 4-5 days, the Th1, Th2 and Tr1 lymphocytes were washed and then expanded again with IL-2 for 4-7 days. Activated Th1 and Th2 lymphocytes were maintained in this way for a maximum of three cycles. RNA was prepared from primary and secondary Th1, Th2 and Tr1 after 6 and 24 hours following the second and third activations with plate bound anti-CD3 and anti-CD28 mAbs and 4 days into the second and third expansion cultures in Interleukin 2.

The following leukocyte cells lines were obtained from the ATCC: Ramos, EOL-1, KU-812. EOL cells were further differentiated by culture in 0.1 mM dbcAMP at 5 x105 cells/ml for 8 days, changing the media every 3 days and adjusting the cell concentration to 5 x105 cells/ml. For the culture of these cells, we used DMEM or RPMI (as recommended by the ATCC), with the addition of 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10-5 M (Gibco), 10 mM Hepes (Gibco). RNA was

either prepared from resting cells or cells activated with PMA at 10 ng/ml and ionomycin at 1 mg/ml for 6 and 14 hours. Keratinocyte line CCD106 and an airway epithelial tumor line NCI-H292 were also obtained from the ATCC. Both were cultured in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10-5 M (Gibco), and 10 mM Hepes (Gibco). CCD1106 cells were activated for 6 and 14 hours with approximately 5 ng/ml TNF alpha and 1 ng/ml IL-1 beta, while NCI-H292 cells were activated for 6 and 14 hours with the following cytokines: 5 ng/ml IL-4, 5 ng/ml IL-9, 5 ng/ml IL-13 and 25 ng/ml IFN gamma.

For these cell lines and blood cells, RNA was prepared by lysing approximately 107 cells/ml using Trizol (Gibco BRL). Briefly, 1/10 volume of bromochloropropane (Molecular Research Corporation) was added to the RNA sample, vortexed and after 10 minutes at room temperature, the tubes were spun at 14,000 rpm in a Sorvall SS34 rotor. The aqueous phase was removed and placed in a 15 ml Falcon Tube. An equal volume of isopropanol was added and left at –20 degrees C overnight. The precipitated RNA was spun down at 9,000 rpm for 15 min in a Sorvall SS34 rotor and washed in 70% ethanol. The pellet was redissolved in 300 ml of RNAse-free water and 35 ml buffer (Promega) 5 ml DTT, 7 ml RNAsin and 8 ml DNAse were added. The tube was incubated at 37 degrees C for 30 minutes to remove contaminating genomic DNA, extracted once with phenol chloroform and re-precipitated with 1/10 volume of 3 M sodium acetate and 2 volumes of 100% ethanol. The RNA was spun down and placed in RNAse free water. RNA was stored at –80 degrees C.

Panel CNSD.01

The plates for Panel CNSD.01 include two control wells and 94 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from the Harvard Brain Tissue Resource Center. Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by neuroanatomists, and frozen at -80oC in liquid nitrogen vapor. All brains are sectioned and examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

Disease diagnoses are taken from patient records. The panel contains two brains from each of the following diagnoses: Alzheimer's disease, Parkinson's disease, Huntington's disease, Progressive Supernuclear Palsy, Depression, and "Normal controls". Within each of these brains, the following regions are represented: cingulate gyrus, temporal pole, globus palladus, substantia nigra, Brodman Area 4 (primary motor strip), Brodman Area 7 (parietal cortex), Brodman Area 9 (prefrontal cortex), and Brodman area 17 (occipital cortex). Not all brain

regions are represented in all cases; e.g., Huntington's disease is characterized in part by neurodegeneration in the globus palladus, thus this region is impossible to obtain from confirmed Huntington's cases. Likewise Parkinson's disease is characterized by degeneration of the substantia nigra making this region more difficult to obtain. Normal control brains were examined for neuropathology and found to be free of any pathology consistent with neurodegeneration.

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

In the labels employed to identify tissues in the CNS panel, the following abbreviations are used:

PSP = Progressive supranuclear palsy

Sub Nigra = Substantia nigra

Glob Palladus= Globus palladus

Temp Pole = Temporal pole

Cing Gyr = Cingulate gyrus

BA 4 = Brodman Area 4

Example 1A - GPCR5 (Gene GM115-a-12-A)

Expression of gene GM115-a-12-A was assessed using the primer-probe set Ag470, described in Table 12. Results of the RTQ-PCR runs are shown in Tables 13, 14, and 15.

Table 12. Probe Name Ag470

Primers	Sequences	SEQ ID	Length
Forward	5'-AAGCAGGGTGCTGTGGTACAG-3'		21
Probe	FAM-5'-TTCTGCGACAGTGGCCCACTGC-3'-TAMRA		22
Reverse	5'-TTCTTGGTGTTGGTGCAAGC-3'		20

Table 13. Panel 1

Tissue Name	Relative Expression(%) 1tm539f_ag470	Tissue Name	Relative Expression(%) 1tm539f_ag470
Endothelial cells	0.0	Kidney (fetal)	0.0
Endothelial cells (treated)	0.0	Renal ca. 786-0	0.0
Pancreas	0.0	Renal ca. A498	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. RXF 393	0.0
Adipose	17.1	Renal ca. ACHN	0.0
Adrenal gland	0.0	Renal ca. UO-31	0.0
Thyroid	0.0	Renal ca. TK-10	0.0
Salivary gland	0.0	Liver	0.0
Pituitary gland	0.0	Liver (fetal)	0.0
Brain (fetal)	0.0	Liver ca. (hepatoblast) HepG2	2.4
Brain (whole)	0.0	Lung	0.0
Brain (amygdala)	0.0	Lung (fetal)	0.0
Brain (cerebellum)	0.0	Lung ca. (small cell) LX-1	0.0
Brain (hippocampus)	0.0	Lung ca. (small cell) NCI-H69	0.7
Brain (substantia nigra)	0.0	Lung ca. (s.cell var.) SHP-77	0.0
Brain (thalamus)	0.0	Lung ca. (large cell)NCI-H460	0.1
Brain (hypothalamus)	0.0	Lung ca. (non-sm. cell) A549	0.0
Spinal cord	0.0	Lung ca. (non-s.cell) NCI-H23	0.0
CNS ca. (glio/astro) U87-MG	0.0	Lung ca (non-s.cell) HOP-62	0.0
CNS ca. (glio/astro) U-118-MG	0.0	Lung ca. (non-s.cl) NCI-H522	0.0
CNS ca. (astro) SW1783	0.0	Lung ca. (squam.) SW 900	0.0
CNS ca.* (neuro; met) SK-N-AS	0.0	Lung ca. (squam.) NCI-H596	0.0
CNS ca. (astro) SF-539	0.0	Mammary gland	0.0
CNS ca. (astro) SNB-75	0.0	Breast ca.* (pl. effusion) MCF-7	0.0
CNS ca. (glio) SNB-19	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.0
CNS ca. (glio) U251	0.0	Breast ca.* (pl. effusion) T47D	0.0
CNS ca. (glio) SF-295	0.0	Breast ca. BT-549	0.0
Heart	0.0	Breast ca. MDA-N	0.0

Skeletal muscle	0.0	Ovary	0.0
Bone marrow	0.0	Ovarian ca. OVCAR-3	0.0
Thymus	0.0	Ovarian ca. OVCAR-4	0.0
Spleen	0.5	Ovarian ca. OVCAR-5	0.9
Lymph node	0.3	Ovarian ca. OVCAR-8	0.0
Colon (ascending)	1.3	Ovarian ca. IGROV-1	0.0
Stomach	0.0	Ovarian ca.* (ascites) SK-OV-3	0.0
Small intestine	0.0	Uterus	**100.0**
Colon ca. SW480	0.0	Placenta	0.0
Colon ca.* (SW480 met)SW620	0.0	Prostate	0.0
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	0.0
Colon ca. HCT-116	0.0	Testis	0.1
Colon ca. CaCo-2	0.0	Melanoma Hs688(A).T	0.0
Colon ca. HCT-15	0.0	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	0.0	Melanoma UACC-62	0.0
Gastric ca.* (liver met) NCI-N87	0.0	Melanoma M14	0.0
Bladder	0.0	Melanoma LOX IMVI	0.0
Trachea	0.0	Melanoma* (met) SK-MEL-5	0.0
Kidney	0.0	Melanoma SK-MEL-28	6.5

Table 14. Panel 1.1

Tissue Name	Relative Expression(%) 1.1tm695f_ ag470	Tissue Name	Relative Expression(%) 1.1tm695f_ ag470
Adipose	**100.0**	Renal ca. TK-10	0.0
Adrenal gland	0.0	Renal ca. UO-31	0.0
Bladder	0.0	Renal ca. RXF 393	0.0
Brain (amygdala)	0.0	Liver	0.0
Brain (cerebellum)	0.0	Liver (fetal)	0.0
Brain (hippocampus)	0.0	Liver ca. (hepatoblast) HepG2	27.9
Brain (substantia nigra)	0.0	Lung	0.0
Brain (thalamus)	0.0	Lung (fetal)	0.0

Cerebral Cortex	0.0	Lung ca (non-s.cell) HOP-62	0.0
Brain (fetal)	0.0	Lung ca. (large cell)NCI-H460	0.0
Brain (whole)	0.0	Lung ca. (non-s.cell) NCI-H23	0.0
CNS ca. (glio/astro) U-118-MG	0.0	Lung ca. (non-s.cl) NCI-H522	0.0
CNS ca. (astro) SF-539	0.0	Lung ca. (non-sm. cell) A549	0.0
CNS ca. (astro) SNB-75	0.0	Lung ca. (s.cell var.) SHP-77	0.0
CNS ca. (astro) SW1783	0.0	Lung ca. (small cell) LX-1	0.0
CNS ca. (glio) U251	0.0	Lung ca. (small cell) NCI-H69	3.2
CNS ca. (glio) SF-295	0.0	Lung ca. (squam.) SW 900	0.0
CNS ca. (glio) SNB-19	0.0	Lung ca. (squam.) NCI-H596	0.0
CNS ca. (glio/astro) U87-MG	0.0	Lymph node	1.0
CNS ca.* (neuro; met) SK-N-AS	0.0	Spleen	0.0
Mammary gland	0.0	Thymus	0.0
Breast ca. BT-549	0.0	Ovary	0.0
Breast ca. MDA-N	0.0	Ovarian ca. IGROV-1	0.0
Breast ca.* (pl. effusion) T47D	0.0	Ovarian ca. OVCAR-3	0.0
Breast ca.* (pl. effusion) MCF-7	0.0	Ovarian ca. OVCAR-4	0.0
Breast ca.* (pl.ef) MDA-MB-231	0.0	Ovarian ca. OVCAR-5	35.1
Small intestine	0.5	Ovarian ca. OVCAR-8	0.0
Colorectal	0.0	Ovarian ca.* (ascites) SK-OV-3	0.0
Colon ca. HT29	0.0	Pancreas	0.0
Colon ca. CaCo-2	0.0	Pancreatic ca. CAPAN 2	0.0
Colon ca. HCT-15	0.0	Pituitary gland	0.0
Colon ca. HCT-116	0.0	Placenta	0.0
Colon ca. HCC-2998	0.0	Prostate	0.0
Colon ca. SW480	0.0	Prostate ca.* (bone met)PC-3	0.0
Colon ca.* (SW480	0.0	Salivary gland	0.0

met)SW620			
Stomach	0.1	Trachea	0.0
Gastric ca.* (liver met) NCI-N87	0.0	Spinal cord	0.0
Heart	0.0	Testis	0.1
Fetal Skeletal	0.0	Thyroid	0.0
Skeletal muscle	0.0	Uterus	0.0
Endothelial cells	0.0	Melanoma M14	0.0
Endothelial cells (treated)	0.0	Melanoma LOX IMVI	0.0
Kidney	0.0	Melanoma UACC-62	0.0
Kidney (fetal)	0.0	Melanoma SK-MEL-28	0.0
Renal ca. 786-0	0.0	Melanoma* (met) SK-MEL-5	0.0
Renal ca. A498	0.0	Melanoma Hs688(A).T	0.0
Renal ca. ACHN	0.0	Melanoma* (met) Hs688(B).T	0.0

Table 15. Panel 1.2

Tissue Name	Relative Expression(%) 1.2tm1 202f_ ag470	Tissue Name	Relative Expression(%) 1.2tm1 202f_ ag470
Endothelial cells	0.0	Renal ca. 786-0	0.0
Endothelial cells (treated)	0.0	Renal ca. A498	0.0
Pancreas	0.0	Renal ca. RXF 393	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. ACHN	0.0
Adrenal Gland (new lot*)	0.0	Renal ca. UO-31	0.0
Thyroid	0.0	Renal ca. TK-10	0.0
Salivary gland	0.0	Liver	0.0
Pituitary gland	0.0	Liver (fetal)	0.0
Brain (fetal)	0.0	Liver ca. (hepatoblast) HepG2	14.4
Brain (whole)	0.0	Lung	0.0
Brain (amygdala)	0.0	Lung (fetal)	0.0
Brain (cerebellum)	0.0	Lung ca. (small cell) LX-1	0.0
Brain (hippocampus)	0.0	Lung ca. (small cell) NCI-H69	1.4

Brain (thalamus)	0.0	Lung ca. (s.cell var.) SHP-77	0.0
Cerebral Cortex	0.0	Lung ca. (large cell)NCI-H460	0.0
Spinal cord	0.0	Lung ca. (non-sm. cell) A549	0.0
CNS ca. (glio/astro) U87-MG	0.0	Lung ca. (non-s.cell) NCI-H23	0.0
CNS ca. (glio/astro) U-118-MG	0.0	Lung ca (non-s.cell) HOP-62	0.0
CNS ca. (astro) SW1783	0.0	Lung ca. (non-s.cl) NCI-H522	0.0
CNS ca.* (neuro; met) SK-N-AS	0.0	Lung ca. (squam.) SW 900	0.0
CNS ca. (astro) SF-539	0.0	Lung ca. (squam.) NCI-H596	0.2
CNS ca. (astro) SNB-75	0.0	Mammary gland	0.0
CNS ca. (glio) SNB-19	0.0	Breast ca.* (pl. effusion) MCF-7	0.0
CNS ca. (glio) U251	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.0
CNS ca. (glio) SF-295	0.0	Breast ca.* (pl. effusion) T47D	0.0
Heart	0.0	Breast ca. BT-549	0.0
Skeletal Muscle (new lot*)	0.0	Breast ca. MDA-N	0.0
Bone marrow	0.0	Ovary	0.0
Thymus	0.0	Ovarian ca. OVCAR-3	0.0
Spleen	0.6	Ovarian ca. OVCAR-4	0.0
Lymph node	0.9	Ovarian ca. OVCAR-5	2.5
Colorectal	0.0	Ovarian ca. OVCAR-8	0.0
Stomach	0.7	Ovarian ca. IGROV-1	0.0
Small intestine	1.5	Ovarian ca.* (ascites) SK-OV-3	0.0
Colon ca. SW480	0.0	Uterus	0.0
Colon ca.* (SW480 met)SW620	0.0	Placenta	0.0
Colon ca. HT29	0.0	Prostate	0.0
Colon ca. HCT-116	0.0	Prostate ca.* (bone met)PC-3	0.0
Colon ca. CaCo-2	0.0	Testis	0.0
83219 CC Well to Mod Diff (ODO3866)	0.6	Melanoma Hs688(A).T	0.0
Colon ca. HCC-2998	0.0	Melanoma* (met) Hs688(B).T	0.0
Gastric ca.* (liver met) NCI-N87	0.0	Melanoma UACC-62	0.0
Bladder	0.0	Melanoma M14	0.0
Trachea	0.0	Melanoma LOX IMVI	0.0

Kidney	0.0	Melanoma* (met) SK-MEL-5	0.0
Kidney (fetal)	0.0	Adipose	**100.0**

Summary of the Panels:

Probe Ag470 for Panel 1 shows that the expression of Nov5a1(the GM115-a-12-A gene) is almost exclusively restricted to the uterus. Thus, the expression of this gene could be used as a unique marker of this tissue.

Probe Ag470 used in Panels 1.1 and 1.2 shows a high level of Nov5a1 (GM115-a-12-A gene) expression in one liver cancer, one ovarian cancer, and one lung cancer cell line, in addition to adipose tissue. As in Panel 1.1, the expression seen in adipose was determined erroneous due to contamination. Thus, therapeutic modulation of this gene may be utilized for the treatment of a subset of ovarian, liver or lung cancers.

Probe Ag470 in Panel 2D resulted in low to undetectable levels of expression of Nov51 (the GM115-a-12-A gene) for all samples on this panel in 3 replicate experiments.

Example 1B - GPCR4 (Gene GMdj408b20C)

Expression of GPCR4 (gene GMdj408b20C) was assessed using the primer-probe set Ag1593, Ag455, and Ag455b, described in Table 12B, 13B, and 14B. Results of the RTQ-PCR runs are shown in Tables 15B, 16, 17, 18, 19, 20, and 21.

Table 12B. Probe Name Ag1593

Primers	Sequences	M	Length	Start Position
Forward	5'-CCTAGAGCTGATCGTCTTTGTG-3'	9.2	22	68
Probe	FAM-5'-TCATCTTTTATCTGCTGACTCTTCTTGGCA-3'-TAMRA	8.7	30	97
Reverse	5'-GAAAGCAAGACAATGGTCATGT-3'	9.1	22	127

Table 13B. Probe Name Ag455

Primers	Sequences	M	Length	Start Position
Forward	5'-CCTGGCTGTCATGGCATATG-3'		20	
Probe	TET-5'-CTATGCTGCAGTCTGCAAACCCCTGC-3'-TAMRA		26	
Reverse	5'-GACGTGGGTGCATGATGATG-3'		20	

Table 14B. Probe Name Ag455b

Primers	Sequences	M	Length	Start Position
Forward	5'-CCTGGCTGTCATGGCATATG-3'		20	
Probe	TET-5'-TGCTGCAGTCTGCAAACCCCTGC-3'-TAMRA		23	
Reverse	5'-GACGTGGGTGCATGATGATG-3'		20	

Table 15B. Panel 1

Tissue Name	Relative Expression(%) 1tm553t_ag455	1tm558t_ag455
Endothelial cells	0.0	0.0
Endothelial cells (treated)	0.1	0.2
Pancreas	0.2	0.2
Pancreatic ca. CAPAN 2	1.0	2.1
Adipose	85.9	**100.0**
Adrenal gland	2.2	3.6
Thyroid	0.0	0.6
Salivary gland	0.5	0.0
Pituitary gland	0.0	0.0
Brain (fetal)	0.6	0.1
Brain (whole)	0.0	0.0
Brain (amygdala)	0.2	0.0
Brain (cerebellum)	0.0	0.0
Brain (hippocampus)	0.0	0.3

Brain (substantia nigra)	0.0	0.5
Brain (thalamus)	0.0	0.5
Brain (hypothalamus)	0.0	0.5
Spinal cord	0.0	0.0
CNS ca. (glio/astro) U87-MG	2.6	4.4
CNS ca. (glio/astro) U-118-MG	7.8	7.2
CNS ca. (astro) SW1783	1.4	3.2
CNS ca.* (neuro; met) SK-N-AS	9.8	10.4
CNS ca. (astro) SF-539	1.7	2.6
CNS ca. (astro) SNB-75	5.2	9.0
CNS ca. (glio) SNB-19	2.4	2.9
CNS ca. (glio) U251	0.7	2.5
CNS ca. (glio) SF-295	7.6	9.0
Heart	0.2	0.0
Skeletal muscle	0.0	0.0
Bone marrow	2.2	2.2
Thymus	9.0	6.8
Spleen	0.4	1.8
Lymph node	0.1	0.0
Colon (ascending)	11.4	19.2
Stomach	0.2	0.4
Small intestine	0.3	0.7
Colon ca. SW480	9.8	11.0
Colon ca.* (SW480 met)SW620	2.6	6.6
Colon ca. HT29	4.0	5.4
Colon ca. HCT-116	0.7	1.4
Colon ca. CaCo-2	7.7	6.2
Colon ca. HCT-15	11.0	14.1
Colon ca. HCC-2998	14.8	16.6
Gastric ca.* (liver met) NCI-N87	55.9	75.8
Bladder	5.0	7.6
Trachea	0.5	0.8
Kidney	0.0	0.0
Kidney (fetal)	0.4	0.8
Renal ca. 786-0	0.3	1.1
Renal ca. A498	4.3	5.6

Renal ca. RXF 393	1.0	0.6
Renal ca. ACHN	0.0	0.3
Renal ca. UO-31	0.7	1.2
Renal ca. TK-10	2.9	4.6
Liver	1.0	1.2
Liver (fetal)	3.2	3.5
Liver ca. (hepatoblast) HepG2	13.4	12.6
Lung	0.5	0.1
Lung (fetal)	0.4	0.1
Lung ca. (small cell) LX-1	2.8	4.2
Lung ca. (small cell) NCI-H69	19.9	15.7
Lung ca. (s.cell var.) SHP-77	6.3	11.0
Lung ca. (large cell)NCI-H460	14.7	29.3
Lung ca. (non-sm. cell) A549	4.3	3.8
Lung ca. (non-s.cell) NCI-H23	18.6	27.4
Lung ca (non-s.cell) HOP-62	4.0	6.7
Lung ca. (non-s.cl) NCI-H522	24.8	32.1
Lung ca. (squam.) SW 900	3.7	4.4
Lung ca. (squam.) NCI-H596	32.8	55.9
Mammary gland	1.5	1.5
Breast ca.* (pl. effusion) MCF-7	**100.0**	86.5
Breast ca.* (pl.ef) MDA-MB-231	6.1	11.7
Breast ca.* (pl. effusion) T47D	11.1	10.5
Breast ca. BT-549	10.2	6.6
Breast ca. MDA-N	15.5	17.8
Ovary	0.0	0.2
Ovarian ca. OVCAR-3	15.7	25.0
Ovarian ca. OVCAR-4	1.8	3.2
Ovarian ca. OVCAR-5	19.6	27.0
Ovarian ca. OVCAR-8	9.9	6.5
Ovarian ca. IGROV-1	5.0	9.2
Ovarian ca.* (ascites) SK-OV-3	1.5	4.0
Uterus	0.2	0.4
Placenta	7.5	8.7
Prostate	2.9	2.3
Prostate ca.* (bone met)PC-3	24.5	33.7

Testis	11.8	10.8
Melanoma Hs688(A).T	0.2	0.4
Melanoma* (met) Hs688(B).T	1.3	2.0
Melanoma UACC-62	0.9	1.0
Melanoma M14	0.4	3.3
Melanoma LOX IMVI	0.0	0.4
Melanoma* (met) SK-MEL-5	1.0	2.4
Melanoma SK-MEL-28	43.8	71.7

Table 16. Panel 1.1

Tissue Name	Relative Expression(%) 1.1tm719t_ag455	Tissue Name	Relative Expression(%) 1.1tm719t_ag455
Adipose	9.1	Renal ca. TK-10	3.8
Adrenal gland	0.0	Renal ca. UO-31	0.6
Bladder	9.0	Renal ca. RXF 393	0.0
Brain (amygdala)	0.0	Liver	0.5
Brain (cerebellum)	0.9	Liver (fetal)	2.3
Brain (hippocampus)	0.0	Liver ca. (hepatoblast) HepG2	7.3
Brain (substantia nigra)	0.8	Lung	0.0
Brain (thalamus)	0.0	Lung (fetal)	1.2
Cerebral Cortex	0.0	Lung ca (non-s.cell) HOP-62	20.6
Brain (fetal)	0.3	Lung ca. (large cell)NCI-H460	7.3
Brain (whole)	0.0	Lung ca. (non-s.cell) NCI-H23	17.1
CNS ca. (glio/astro) U-118-MG	6.9	Lung ca. (non-s.cl) NCI-H522	52.1
CNS ca. (astro) SF-539	0.0	Lung ca. (non-sm. cell) A549	4.8
CNS ca. (astro) SNB-75	7.9	Lung ca. (s.cell var.) SHP-77	12.9

CNS ca. (astro) SW1783	0.6	Lung ca. (small cell) LX-1	6.8
CNS ca. (glio) U251	2.1	Lung ca. (small cell) NCI-H69	17.0
CNS ca. (glio) SF-295	13.3	Lung ca. (squam.) SW 900	2.5
CNS ca. (glio) SNB-19	3.0	Lung ca. (squam.) NCI-H596	54.3
CNS ca. (glio/astro) U87-MG	3.8	Lymph node	0.3
CNS ca.* (neuro; met) SK-N-AS	9.4	Spleen	0.0
Mammary gland	0.0	Thymus	2.7
Breast ca. BT-549	6.2	Ovary	0.0
Breast ca. MDA-N	13.8	Ovarian ca. IGROV-1	8.2
Breast ca.* (pl. effusion) T47D	13.2	Ovarian ca. OVCAR-3	30.8
Breast ca.* (pl. effusion) MCF-7	**100.0**	Ovarian ca. OVCAR-4	3.4
Breast ca.* (pl.ef) MDA-MB-231	11.5	Ovarian ca. OVCAR-5	32.5
Small intestine	1.7	Ovarian ca. OVCAR-8	4.3
Colorectal	0.8	Ovarian ca.* (ascites) SK-OV-3	4.3
Colon ca. HT29	4.0	Pancreas	1.0
Colon ca. CaCo-2	0.3	Pancreatic ca. CAPAN 2	1.1
Colon ca. HCT-15	7.3	Pituitary gland	0.0
Colon ca. HCT-116	2.2	Placenta	12.3
Colon ca. HCC-2998	15.6	Prostate	1.9
Colon ca. SW480	7.9	Prostate ca.* (bone met)PC-3	24.1
Colon ca.* (SW480 met)SW620	13.5	Salivary gland	0.0

Stomach	1.0	Trachea	0.1
Gastric ca.* (liver met) NCI-N87	88.9	Spinal cord	0.0
Heart	0.2	Testis	6.4
Fetal Skeletal	0.7	Thyroid	0.2
Skeletal muscle	0.0	Uterus	0.0
Endothelial cells	0.8	Melanoma M14	3.8
Endothelial cells (treated)	0.3	Melanoma LOX IMVI	0.7
Kidney	0.0	Melanoma UACC-62	1.1
Kidney (fetal)	0.6	Melanoma SK-MEL-28	5.9
Renal ca. 786-0	0.9	Melanoma* (met) SK-MEL-5	0.9
Renal ca. A498	6.1	Melanoma Hs688(A).T	0.2
enal ca. ACHN	0.9	Melanoma* (met) Hs688(B).T	1.5

Table 17. Panel 1.2

Tissue Name	Relative Expression(%) 1.2tm1187t_ ag455b	Tissue Name	Relative Expression(%) 1.2tm1187t_ ag455b
Endothelial cells	0.0	Renal ca. 786-0	0.0
Endothelial cells (treated)	0.0	Renal ca. A498	4.9
Pancreas	0.0	Renal ca. RXF 393	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. ACHN	0.0
Adrenal Gland (new lot*)	0.2	Renal ca. UO-31	0.0
Thyroid	0.0	Renal ca. TK-10	1.1
Salivary gland	0.0	Liver	0.5
Pituitary gland	0.0	Liver (fetal)	1.6
Brain (fetal)	0.0	Liver ca. (hepatoblast) HepG2	7.6
Brain (whole)	0.0	Lung	0.0
Brain (amygdala)	0.0	Lung (fetal)	0.4

Brain (cerebellum)	0.1	Lung ca. (small cell) LX-1	1.8
Brain (hippocampus)	0.0	Lung ca. (small cell) NCI-H69	9.9
Brain (thalamus)	0.0	Lung ca. (s.cell var.) SHP-77	8.4
Cerebral Cortex	0.0	Lung ca. (large cell)NCI-H460	13.3
Spinal cord	0.0	Lung ca. (non-sm. cell) A549	1.9
CNS ca. (glio/astro) U87-MG	2.0	Lung ca. (non-s.cell) NCI-H23	19.3
CNS ca. (glio/astro) U-118-MG	3.0	Lung ca (non-s.cell) HOP-62	0.7
CNS ca. (astro) SW1783	0.3	Lung ca. (non-s.cl) NCI-H522	29.7
CNS ca.* (neuro; met) SK-N-AS	5.3	Lung ca. (squam.) SW 900	1.7
CNS ca. (astro) SF-539	0.0	Lung ca. (squam.) NCI-H596	31.0
CNS ca. (astro) SNB-75	8.0	Mammary gland	0.0
CNS ca. (glio) SNB-19	0.3	Breast ca.* (pl. effusion) MCF-7	**100.0**
CNS ca. (glio) U251	1.3	Breast ca.* (pl.ef) MDA-MB-231	5.7
CNS ca. (glio) SF-295	2.7	Breast ca.* (pl. effusion) T47D	7.5
Heart	0.0	Breast ca. BT-549	18.7
Skeletal Muscle (new lot*)	0.0	Breast ca. MDA-N	7.6
Bone marrow	0.7	Ovary	0.0
Thymus	0.8	Ovarian ca. OVCAR-3	11.8
Spleen	0.0	Ovarian ca. OVCAR-4	1.3
Lymph node	0.0	Ovarian ca. OVCAR-5	16.6
Colorectal	0.6	Ovarian ca. OVCAR-8	1.2
Stomach	0.6	Ovarian ca. IGROV-1	2.5
Small intestine	0.0	Ovarian ca.* (ascites) SK-OV-3	1.1
Colon ca. SW480	4.1	Uterus	0.0
Colon ca.* (SW480 met)SW620	3.8	Placenta	19.1
Colon ca. HT29	0.4	Prostate	2.0
Colon ca. HCT-116	0.6	Prostate ca.* (bone met)PC-3	0.0
Colon ca. CaCo-2	2.4	Testis	10.7
83219 CC Well to Mod Diff (ODO3866)	4.4	Melanoma Hs688(A).T	0.0
Colon ca. HCC-2998	9.7	Melanoma* (met) Hs688(B).T	0.0
Gastric ca.* (liver met) NCI-N87	49.3	Melanoma UACC-62	0.0

Bladder	4.9	Melanoma M14	0.1
Trachea	0.0	Melanoma LOX IMVI	0.0
Kidney	0.0	Melanoma* (met) SK-MEL-5	0.2
Kidney (fetal)	0.5	Adipose	59.9

Table 18. Panel 1.3D

Tissue Name	Relative Expression(%) 1.3Dtm3265f_ ag1593	Relative Expression(%) 1.3dx4tm5603t _ag455b_b2
Liver adenocarcinoma	0.0	0.0
Pancreas	0.0	5.1
Pancreatic ca. CAPAN 2	0.0	0.0
Adrenal gland	0.0	0.0
Thyroid	0.0	0.0
Salivary gland	0.0	0.0
Pituitary gland	0.0	0.0
Brain (fetal)	0.0	0.0
Brain (whole)	0.0	0.0
Brain (amygdala)	0.0	0.0
Brain (cerebellum)	0.0	0.0
Brain (hippocampus)	0.0	0.0
Brain (substantia nigra)	0.0	0.0
Brain (thalamus)	0.0	0.0
Cerebral Cortex	0.0	0.0
Spinal cord	0.0	0.0
CNS ca. (glio/astro) U87-MG	2.1	0.0
CNS ca. (glio/astro) U-118-MG	27.7	14.4
CNS ca. (astro) SW1783	6.0	0.0
CNS ca.* (neuro; met) SK-N-AS	20.4	0.0
CNS ca. (astro) SF-539	0.0	0.0
CNS ca. (astro) SNB-75	6.3	10.5
CNS ca. (glio) SNB-19	1.7	0.0
CNS ca. (glio) U251	0.0	0.0
CNS ca. (glio) SF-295	14.3	6.2

Heart (fetal)	2.4	0.0
Heart	0.0	0.0
Fetal Skeletal	3.8	4.7
Skeletal muscle	0.0	0.0
Bone marrow	0.0	0.0
Thymus	0.0	3.3
Spleen	0.0	0.0
Lymph node	0.0	0.0
Colorectal	5.6	4.5
Stomach	0.0	0.0
Small intestine	2.0	0.0
Colon ca. SW480	22.4	8.5
Colon ca.* (SW480 met)SW620	8.4	10.8
Colon ca. HT29	8.2	0.0
Colon ca. HCT-116	1.7	6.5
Colon ca. CaCo-2	7.4	0.0
83219 CC Well to Mod Diff (ODO3866)	3.7	0.0
Colon ca. HCC-2998	9.1	23.1
Gastric ca.* (liver met) NCI-N87	64.6	43.5
Bladder	0.0	0.0
Trachea	0.0	6.3
Kidney	0.0	0.0
Kidney (fetal)	2.8	0.0
Renal ca. 786-0	3.6	0.0
Renal ca. A498	8.5	0.0
Renal ca. RXF 393	1.8	0.0
Renal ca. ACHN	2.5	0.7
Renal ca. UO-31	0.0	0.0
Renal ca. TK-10	0.0	0.0
Liver	0.0	5.6
Liver (fetal)	0.0	0.0
Liver ca. (hepatoblast) HepG2	8.7	3.9
Lung	0.0	0.0
Lung (fetal)	2.0	0.0
Lung ca. (small cell) LX-1	4.2	0.0
Lung ca. (small cell) NCI-H69	11.9	17.1

Lung ca. (s.cell var.) SHP-77	28.1	33.4
Lung ca. (large cell)NCI-H460	3.8	0.0
Lung ca. (non-sm. cell) A549	1.9	0.0
Lung ca. (non-s.cell) NCI-H23	24.7	17.1
Lung ca (non-s.cell) HOP-62	5.8	4.1
Lung ca. (non-s.cl) NCI-H522	25.3	0.0
Lung ca. (squam.) SW 900	0.0	5.0
Lung ca. (squam.) NCI-H596	32.8	31.7
Mammary gland	0.0	0.0
Breast ca.* (pl. effusion) MCF-7	**100.0**	**100.0**
Breast ca.* (pl.ef) MDA-MB-231	25.7	29.1
Breast ca.* (pl. effusion) T47D	8.5	5.8
Breast ca. BT-549	26.6	11.3
Breast ca. MDA-N	6.7	0.0
Ovary	0.0	0.0
Ovarian ca. OVCAR-3	6.5	17.3
Ovarian ca. OVCAR-4	3.6	0.0
Ovarian ca. OVCAR-5	6.1	9.8
Ovarian ca. OVCAR-8	2.0	0.0
Ovarian ca. IGROV-1	0.0	8.1
Ovarian ca.* (ascites) SK-OV-3	2.7	0.0
Uterus	3.6	0.0
Placenta	4.1	27.0
Prostate	0.0	5.2
Prostate ca.* (bone met)PC-3	6.9	0.0
Testis	9.3	17.1
Melanoma Hs688(A).T	0.0	0.0
Melanoma* (met) Hs688(B).T	0.0	0.0
Melanoma UACC-62	1.6	0.0
Melanoma M14	0.0	0.0
Melanoma LOX IMVI	2.4	0.0
Melanoma* (met) SK-MEL-5	1.9	0.0
Adipose	0.0	0.0

Table 19. Panel 2D

Tissue Name	Relative Expression(%) 2Dtm3266f_ ag1593	Relative Expression(%) 2Dtm2398t_ ag455	2dtm2415t_ ag455
Normal Colon GENPAK 061003	2.5	0.5	2.5
83219 CC Well to Mod Diff (ODO3866)	1.1	0.0	0.6
83220 CC NAT (ODO3866)	2.1	1.2	1.0
83221 CC Gr.2 rectosigmoid (ODO3868)	3.2	1.3	3.2
83222 CC NAT (ODO3868)	0.0	0.0	0.0
83235 CC Mod Diff (ODO3920)	8.7	3.0	2.4
83236 CC NAT (ODO3920)	1.9	0.0	0.0
83237 CC Gr.2 ascend colon (ODO3921)	3.0	1.7	1.4
83238 CC NAT (ODO3921)	0.5	0.4	0.0
83241 CC from Partial Hepatectomy (ODO4309)	1.7	2.6	0.0
83242 Liver NAT (ODO4309)	0.5	1.2	0.6
87472 Colon mets to lung (OD04451-01)	0.0	0.0	0.5
87473 Lung NAT (OD04451-02)	0.9	0.0	0.0
Normal Prostate Clontech A+ 6546-1	0.9	0.5	1.4
84140 Prostate Cancer (OD04410)	16.3	11.9	2.9
84141 Prostate NAT (OD04410)	5.6	3.8	0.9
87073 Prostate Cancer (OD04720-01)	0.5	1.1	0.0
87074 Prostate NAT (OD04720-02)	0.5	1.8	0.0
Normal Lung GENPAK 061010	2.8	1.5	0.0
83239 Lung Met to Muscle (ODO4286)	3.6	10.4	1.1
83240 Muscle NAT (ODO4286)	0.0	0.0	0.7
84136 Lung Malignant Cancer (OD03126)	0.0	1.8	0.7
84137 Lung NAT (OD03126)	0.0	1.4	0.7
84871 Lung Cancer (OD04404)	2.2	0.7	2.5
84872 Lung NAT (OD04404)	0.0	0.0	1.5
84875 Lung Cancer (OD04565)	0.4	0.4	1.9
84876 Lung NAT (OD04565)	0.0	0.0	0.0
85950 Lung Cancer (OD04237-01)	7.6	5.4	5.5
85970 Lung NAT (OD04237-02)	0.4	0.4	0.0
83255 Ocular Mel Met to Liver (ODO4310)	0.0	0.0	0.7
83256 Liver NAT (ODO4310)	0.0	0.0	0.0

84139 Melanoma Mets to Lung (OD04321)	3.6	6.8	0.8
84138 Lung NAT (OD04321)	1.0	3.1	0.0
Normal Kidney GENPAK 061008	0.0	0.4	0.0
83786 Kidney Ca, Nuclear grade 2 (OD04338)	0.9	0.0	2.3
83787 Kidney NAT (OD04338)	1.2	0.8	0.8
83788 Kidney Ca Nuclear grade 1/2 (OD04339)	1.9	0.7	1.5
83789 Kidney NAT (OD04339)	0.0	0.0	0.0
83790 Kidney Ca, Clear cell type (OD04340)	0.4	0.0	0.0
83791 Kidney NAT (OD04340)	0.8	0.0	0.9
83792 Kidney Ca, Nuclear grade 3 (OD04348)	0.5	3.3	0.0
83793 Kidney NAT (OD04348)	2.0	2.4	3.0
87474 Kidney Cancer (OD04622-01)	0.0	1.7	0.0
87475 Kidney NAT (OD04622-03)	0.4	0.9	0.0
85973 Kidney Cancer (OD04450-01)	0.0	0.4	0.0
85974 Kidney NAT (OD04450-03)	0.0	0.0	0.0
Kidney Cancer Clontech 8120607	0.0	0.0	0.0
Kidney NAT Clontech 8120608	0.0	0.0	0.0
Kidney Cancer Clontech 8120613	0.0	0.0	0.0
Kidney NAT Clontech 8120614	0.0	0.0	0.0
Kidney Cancer Clontech 9010320	0.9	0.0	2.5
Kidney NAT Clontech 9010321	0.0	0.0	1.0
Normal Uterus GENPAK 061018	0.0	0.0	0.6
Uterus Cancer GENPAK 064011	0.4	0.4	0.0
Normal Thyroid Clontech A+ 6570-1	0.0	0.5	0.6
Thyroid Cancer GENPAK 064010	0.4	1.5	0.9
Thyroid Cancer INVITROGEN A302152	3.0	2.9	0.8
Thyroid NAT INVITROGEN A302153	0.0	0.8	0.5
Normal Breast GENPAK 061019	0.0	0.3	2.1
84877 Breast Cancer (OD04566)	6.2	7.1	4.9
85975 Breast Cancer (OD04590-01)	**100.0**	49.3	57.0
85976 Breast Cancer Mets (OD04590-03)	37.4	**100.0**	40.3
87070 Breast Cancer Metastasis (OD04655-05)	13.8	27.4	11.0
GENPAK Breast Cancer 064006	3.3	5.1	0.8
Breast Cancer Res. Gen. 1024	8.2	8.1	13.1
Breast Cancer Clontech 9100266	1.0	3.0	1.4
Breast NAT Clontech 9100265	1.7	0.0	1.7

Breast Cancer INVITROGEN A209073	1.5	3.2	0.7
Breast NAT INVITROGEN A2090734	0.4	0.8	1.4
Normal Liver GENPAK 061009	0.0	2.9	0.0
Liver Cancer GENPAK 064003	0.0	0.4	0.9
Liver Cancer Research Genetics RNA 1025	2.1	0.4	0.7
Liver Cancer Research Genetics RNA 1026	0.0	0.0	0.0
Paired Liver Cancer Tissue Research Genetics RNA 6004-T	0.4	0.0	0.0
Paired Liver Tissue Research Genetics RNA 6004-N	4.2	0.7	2.9
Paired Liver Cancer Tissue Research Genetics RNA 6005-T	0.7	0.0	2.4
Paired Liver Tissue Research Genetics RNA 6005-N	0.0	0.4	0.0
Normal Bladder GENPAK 061001	3.6	5.4	0.8
Bladder Cancer Research Genetics RNA 1023	0.7	0.5	1.8
Bladder Cancer INVITROGEN A302173	4.5	2.1	5.4
87071 Bladder Cancer (OD04718-01)	33.4	64.2	20.4
87072 Bladder Normal Adjacent (OD04718-03)	0.0	0.7	0.0
Normal Ovary Res. Gen.	0.0	0.0	0.0
Ovarian Cancer GENPAK 064008	2.6	2.4	0.7
87492 Ovary Cancer (OD04768-07)	97.9	65.1	**100.0**
87493 Ovary NAT (OD04768-08)	0.0	0.5	0.0
Normal Stomach GENPAK 061017	0.0	1.2	1.5
Gastric Cancer Clontech 9060358	0.0	0.0	2.2
NAT Stomach Clontech 9060359	0.4	0.0	0.0
Gastric Cancer Clontech 9060395	0.0	0.0	0.0
NAT Stomach Clontech 9060394	0.0	1.0	0.0
Gastric Cancer Clontech 9060397	4.1	2.3	1.0
NAT Stomach Clontech 9060396	0.0	1.3	1.1
Gastric Cancer GENPAK 064005	2.9	0.4	1.6

Table 20. Panel 3D

Tissue Name	Relative Expression(%)	Tissue Name	Relative Expression(%)

	3dx4tm5097t_ag455_a2		3dx4tm5097t_ag455_a2
94905_Daoy_Medulloblastoma/Cerebellum_sscDNA	18.2	94954_Ca Ski_Cervical epidermoid carcinoma (metastasis)_sscDNA	33.9
94906_TE671_Medulloblastom/Cerebellum_sscDNA	0.8	94955_ES-2_Ovarian clear cell carcinoma_sscDNA	0.0
94907_D283 Med_Medulloblastoma/Cerebellum_sscDNA	3.8	94957_Ramos/6h stim_"; Stimulated with PMA/ionomycin 6h_sscDNA	1.2
94908_PFSK-1_Primitive Neuroectodermal/Cerebellum_sscDNA	4.5	94958_Ramos/14h stim_"; Stimulated with PMA/ionomycin 14h_sscDNA	1.5
94909_XF-498_CNS_sscDNA	4.5	4962_MEG-01_Chronic myelogenous leukemia (megokaryoblast)_sscDNA	**100.0**
94910_SNB-78_CNS/glioma_sscDNA	2.0	94963_Raji_Burkitt's lymphoma_sscDNA	0.0
94911_SF-268_CNS/glioblastoma_sscDNA	0.0	94964_Daudi_Burkitt's lymphoma_sscDNA	0.0
94912_T98G_Glioblastoma_sscDNA	7.9	94965_U266_B-cell plasmacytoma/myeloma_sscDNA	40.2
96776_SK-N-SH_Neuroblastoma (metastasis)_sscDNA	0.0	94968_CA46_Burkitt's lymphoma_sscDNA	1.6
94913_SF-295_CNS/glioblastoma_sscDNA	1.9	94970_RL_non-Hodgkin's B-cell lymphoma_sscDNA	0.0
94914_Cerebellum_sscDNA	1.2	94972_JM1_pre-B-cell lymphoma/leukemia_sscDNA	3.7
96777_Cerebellum_sscDNA	2.1	94973_Jurkat_T cell leukemia_sscDNA	0.0
94916_NCI-H292_Mucoepidermoid lung carcinoma_sscDNA	2.4	94974_TF-1_Erythroleukemia_sscDNA	10.6

94917_DMS-114_Small cell lung cancer_sscDNA	0.2	94975_HUT 78_T-cell lymphoma_sscDNA	2.9
94918_DMS-79_Small cell lung cancer/neuroendocrine_sscDNA	98.0	94977_U937_Histiocytic lymphoma_sscDNA	12.2
94919_NCI-H146_Small cell lung cancer/neuroendocrine_sscDNA	22.6	94980_KU-812_Myelogenous leukemia_sscDNA	17.6
94920_NCI-H526_Small cell lung cancer/neuroendocrine_sscDNA	2.3	94981_769-P_Clear cell renal carcinoma_sscDNA	0.0
94921_NCI-N417_Small cell lung cancer/neuroendocrine_sscDNA	14.4	94983_Caki-2_Clear cell renal carcinoma_sscDNA	0.0
94923_NCI-H82_Small cell lung cancer/neuroendocrine_sscDNA	3.6	94984_SW 839_Clear cell renal carcinoma_sscDNA	1.6
94924_NCI-H157_Squamous cell lung cancer (metastasis)_sscDNA	34.1	94986_G401_Wilms' tumor_sscDNA	0.0
94925_NCI-H1155_Large cell lung cancer/neuroendocrine_sscDNA	9.8	94987_Hs766T_Pancreatic carcinoma (LN metastasis)_sscDNA	11.6
94926_NCI-H1299_Large cell lung cancer/neuroendocrine_sscDNA	18.2	94988_CAPAN-1_Pancreatic adenocarcinoma (liver metastasis)_sscDNA	6.6
94927_NCI-H727_Lung carcinoid_sscDNA	1.8	94989_SU86.86_Pancreatic carcinoma (liver metastasis)_sscDNA	3.0
94928_NCI-UMC-11_Lung carcinoid_sscDNA	7.9	94990_BxPC-3_Pancreatic adenocarcinoma_sscDNA	13.3
94929_LX-1_Small cell lung cancer_sscDNA	1.3	94991_HPAC_Pancreatic adenocarcinoma_sscDNA	2.8
94930_Colo-205_Colon cancer_sscDNA	0.0	94992_MIA PaCa-2_Pancreatic carcinoma_sscDNA	0.0

94931_KM12_Colon cancer_sscDNA	23.6	94993_CFPAC-1_Pancreatic ductal adenocarcinoma_sscDNA	17.6
94932_KM20L2_Colon cancer_sscDNA	0.0	94994_PANC-1_Pancreatic epithelioid ductal carcinoma_sscDNA	5.5
94933_NCI-H716_Colon cancer_sscDNA	40.5	94996_T24_Bladder carcinma (transitional cell)_sscDNA	5.0
94935_SW-48_Colon adenocarcinoma_sscDNA	1.8	94997_5637_Bladder carcinoma_sscDNA	0.0
94936_SW1116_Colon adenocarcinoma_sscDNA	2.3	94998_HT-1197_Bladder carcinoma_sscDNA	15.1
94937_LS 174T_Colon adenocarcinoma_sscDNA	0.0	94999_UM-UC-3_Bladder carcinma (transitional cell)_sscDNA	0.0
94938_SW-948_Colon adenocarcinoma_sscDNA	1.1	95000_A204_Rhabdomyosarcoma_sscDNA	5.6
94939_SW-480_Colon adenocarcinoma_sscDNA	0.0	95001_HT-1080_Fibrosarcoma_sscDNA	0.0
94940_NCI-SNU-5_Gastric carcinoma_sscDNA	11.0	95002_MG-63_Osteosarcoma (bone)_sscDNA	2.0
94941_KATO III_Gastric carcinoma_sscDNA	9.7	95003_SK-LMS-1_Leiomyosarcoma (vulva)_sscDNA	10.7
94943_NCI-SNU-16_Gastric carcinoma_sscDNA	11.6	95004_SJRH30_Rhabdomyosarcoma (met to bone marrow)_sscDNA	1.6
94944_NCI-SNU-1_Gastric carcinoma_sscDNA	6.0	95005_A431_Epidermoid carcinoma_sscDNA	4.6
94946_RF-1_Gastric adenocarcinoma_sscDNA	2.6	95007_WM266-4_Melanoma_sscDNA	1.3
94947_RF-48_Gastric adenocarcinoma_sscDNA	6.0	95010_DU 145_Prostate carcinoma (brain metastasis)_sscDNA	0.0

96778_MKN-45_Gastric carcinoma_sscDNA	0.0	95012_MDA-MB-468_Breast adenocarcinoma_sscDNA	8.5
94949_NCI-N87_Gastric carcinoma_sscDNA	1.4	95013_SCC-4_Squamous cell carcinoma of tongue_sscDNA	0.0
94951_OVCAR-5_Ovarian carcinoma_sscDNA	0.0	95014_SCC-9_Squamous cell carcinoma of tongue_sscDNA	1.2
94952_RL95-2_Uterine carcinoma_sscDNA	3.7	95015_SCC-15_Squamous cell carcinoma of tongue_sscDNA	1.8
94953_HelaS3_Cervical adenocarcinoma_sscDNA	30.0	95017_CAL 27_Squamous cell carcinoma of tongue_sscDNA	0.0

Table 21. Panel 4D

Tissue Name	Relative Expression(%) 4Dtm3267f_ag1593	Relative Expression(%) 4Dtm1764t_ag455
93768_Secondary Th1_anti-CD28/anti-CD3	7.5	4.0
93769_Secondary Th2_anti-CD28/anti-CD3	9.9	5.5
93770_Secondary Tr1_anti-CD28/anti-CD3	5.0	7.7
93573_Secondary Th1_resting day 4-6 in IL-2	7.2	5.1
93572_Secondary Th2_resting day 4-6 in IL-2	5.3	4.6
93571_Secondary Tr1_resting day 4-6 in IL-2	0.0	8.1
93568_primary Th1_anti-CD28/anti-CD3	0.0	6.6
93569_primary Th2_anti-CD28/anti-CD3	5.1	0.0
93570_primary Tr1_anti-CD28/anti-CD3	0.0	1.5
93565_primary Th1_resting dy 4-6 in IL-2	14.8	6.4
93566_primary Th2_resting dy 4-6 in IL-2	4.3	4.1
93567_primary Tr1_resting dy 4-6 in IL-2	4.4	0.9
93351_CD45RA CD4 lymphocyte_anti-CD28/anti-CD3	6.2	10.4
93352_CD45RO CD4 lymphocyte_anti-CD28/anti-CD3	1.6	3.8

93251_CD8 Lymphocytes_anti-CD28/anti-CD3	0.0	1.5
93353_chronic CD8 Lymphocytes 2ry_resting dy 4-6 in IL-2	0.0	1.5
93574_chronic CD8 Lymphocytes 2ry_activated CD3/CD28	0.0	5.3
93354_CD4_none	0.0	0.0
93252_Secondary Th1/Th2/Tr1_anti-CD95 CH11	8.8	5.1
93103_LAK cells_resting	0.0	3.1
93788_LAK cells_IL-2	5.2	14.0
93787_LAK cells_IL-2+IL-12	1.6	6.7
93789_LAK cells_IL-2+IFN gamma	8.9	6.7
93790_LAK cells_IL-2+ IL-18	7.0	1.2
93104_LAK cells_PMA/ionomycin and IL-18	1.6	0.0
93578_NK Cells IL-2_resting	3.8	5.2
93109_Mixed Lymphocyte Reaction_Two Way MLR	6.3	4.3
93110_Mixed Lymphocyte Reaction_Two Way MLR	1.7	3.4
93111_Mixed Lymphocyte Reaction_Two Way MLR	0.0	1.7
93112_Mononuclear Cells (PBMCs)_resting	0.0	1.3
93113_Mononuclear Cells (PBMCs)_PWM	28.1	25.7
93114_Mononuclear Cells (PBMCs)_PHA-L	1.1	10.1
93249_Ramos (B cell)_none	6.7	3.6
93250_Ramos (B cell)_ionomycin	9.1	16.0
93349_B lymphocytes_PWM	37.6	47.6
93350_B lymphoytes_CD40L and IL-4	6.4	14.4
92665_EOL-1 (Eosinophil)_dbcAMP differentiated	0.0	2.8
93248_EOL-1 (Eosinophil)_dbcAMP/PMAionomycin	1.2	1.8
93356_Dendritic Cells_none	0.0	0.0
93355_Dendritic Cells_LPS 100 ng/ml	0.0	0.0
93775_Dendritic Cells_anti-CD40	0.0	0.0
93774_Monocytes_resting	0.0	0.0
93776_Monocytes_LPS 50 ng/ml	0.0	4.7
93581_Macrophages_resting	0.0	0.0
93582_Macrophages_LPS 100 ng/ml	0.0	1.7
93098_HUVEC (Endothelial)_none	3.0	0.0
93099_HUVEC (Endothelial)_starved	3.1	3.3
93100_HUVEC (Endothelial)_IL-1b	0.0	0.0
93779_HUVEC (Endothelial)_IFN gamma	6.3	0.0
93102_HUVEC (Endothelial)_TNF alpha + IFN gamma	0.0	0.0

93101_HUVEC (Endothelial)_TNF alpha + IL4	0.0	0.0
93781_HUVEC (Endothelial)_IL-11	0.0	3.4
93583_Lung Microvascular Endothelial Cells_none	0.0	0.0
93584_Lung Microvascular Endothelial Cells_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	1.6
92662_Microvascular Dermal endothelium_none	0.0	3.6
92663_Microsvasular Dermal endothelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	7.1	0.0
93773_Bronchial epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml) **	0.0	0.0
93347_Small Airway Epithelium_none	0.0	3.8
93348_Small Airway Epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	2.3
92668_Coronery Artery SMC_resting	0.0	1.6
92669_Coronery Artery SMC_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	6.4
93107_astrocytes_resting	6.8	0.0
93108_astrocytes_TNFa (4 ng/ml) and IL1b (1 ng/ml)	1.7	6.0
92666_KU-812 (Basophil)_resting	13.7	25.9
92667_KU-812 (Basophil)_PMA/ionoycin	**100.0**	**100.0**
93579_CCD1106 (Keratinocytes)_none	4.5	2.1
93580_CCD1106 (Keratinocytes)_TNFa and IFNg **	0.0	12.1
93791_Liver Cirrhosis	3.5	14.9
93792_Lupus Kidney	0.0	0.0
93577_NCI-H292	2.5	3.3
93358_NCI-H292_IL-4	13.8	7.5
93360_NCI-H292_IL-9	9.7	7.9
93359_NCI-H292_IL-13	1.7	5.4
93357_NCI-H292_IFN gamma	6.8	10.9
93777_HPAEC_-	0.0	0.0
93778_HPAEC_IL-1 beta/TNA alpha	9.6	3.3
93254_Normal Human Lung Fibroblast_none	3.0	0.0
93253_Normal Human Lung Fibroblast_TNFa (4 ng/ml) and IL-1b (1 ng/ml)	6.8	0.0
93257_Normal Human Lung Fibroblast_IL-4	1.5	0.0
93256_Normal Human Lung Fibroblast_IL-9	2.9	0.0

93255_Normal Human Lung Fibroblast_IL-13	0.0	2.6
93258_Normal Human Lung Fibroblast_IFN gamma	0.0	2.5
93106_Dermal Fibroblasts CCD1070_resting	2.5	12.4
93361_Dermal Fibroblasts CCD1070_TNF alpha 4 ng/ml	4.9	12.4
93105_Dermal Fibroblasts CCD1070_IL-1 beta 1 ng/ml	0.0	5.2
93772_dermal fibroblast_IFN gamma	0.0	0.0
93771_dermal fibroblast_IL-4	1.6	4.3
93259_IBD Colitis 1**	0.0	31.0
93260_IBD Colitis 2	3.4	1.4
93261_IBD Crohns	1.5	1.6
735010_Colon_normal	6.9	2.0
735019_Lung_none	2.2	0.0
64028-1_Thymus_none	0.0	0.0
64030-1_Kidney_none	36.6	62.0

Summary of Panels:

Probe Ag455 in Panel 1 indicate that the results from two replicate experiments using the same probe and primer set are in good agreement. GPCR4 (the GMdj408b20C gene) appears to be predominantly expressed in cell lines rather than normal tissues. There is high expression in MCF-7 breast cancer cells, which are estrogen receptor positive. This suggests that the expression of the gene is associated with breast cancer cells or with a relatively large fraction of cells that are actively cycling. What appears to be high expression in adipose tissue is due to contamination with genomic DNA.

Probe Ag455 in Panel 1.1 is similar to that observed in Panel 1 for the overall expression pattern of GPCR4 (the GMdj408b20C gene). Specifically, this gene appears to be over expressed in breast, ovarian, prostate, lung, gastric and CNS cancer cell lines relative to normal controls. Therefore, monoclonal antibodies and/or small molecule therapeutics designed to inhibit the GMdj408b20C gene product may have utility in the treatment of these cancers.

Probe Ag455b in Panel 1.2 also shows that GPCR4 (the GMdj408b20C gene) is over expressed in breast, ovarian, lung, gastric and CNS cancer cell lines relative to normal controls. The highest level of expression was found in MCF-7 breast cancer cells, which are estrogen receptor positive. As with Panel 1, this suggests an association of the gene with cancer cells, particularly breast cancer, or a cell population in which a relatively large fraction of cells are actively cycling. Again, the apparent high level of expression in adipose tissue is due to contamination.

Probes Ag1593/Ag455b in Panel 1.3D evidence good agreement from two experiments in which different probes and primer sets were utilized. Expression of the GMdj408b20C gene appears to be clustered to cell lines rather than to normal tissues. The highest expression was found in MCF-7 breast cancer cells, which are estrogen receptor positive. This suggests an association with cancer cells, particularly breast cancer cells, or cell populations that are actively cycling.

Probes Ag1593/Ag455 in panel 2D shows that expression of the GMdj408b20C gene across these samples is low. However, consistent expression in breast cancers is found, and some over expression is seen in ovarian and bladder cancers relative to adjacent normal tissue. Thus, therapeutic modulation of this gene may be beneficial for treatment of these cancers. The results in Panel 2D are consistent with what was observed in Panels 1, 1.1, 1.2 and 1.3D.

Probe Ag455 in Panel 3D shows a wide but low level of expression for the GMdj408b20C gene across all samples in the panel. Highest expression was detected in a chronic myelogenous leukemia (megokaryoblast), which suggests a potential role for the gene in this disease.

Probe Ag1593/Ag455 in Panel 4D indicates good agreement between results from two experiments using different probe and primer sets. The GMdj408b20C gene is highly induced the KU-812 basophil cell line, but is expressed at lower levels in pokeweed mitogen-activated B cells and PBMC. It is known that activated basophils release a number of potent bioresponse modifiers capable of damaging surrounding tissues. Antibody or small molecule therapies directed against the protein encoded for by the GMdj408b20C gene may reduce or block inflammation and/or tissue damage caused by inflammation through blockade of basophil activation. Such therapies would be important for the treatment of asthma, emphysema, and allergy.

Probes ag1593/ag455 in panel cnsd.01 show low to undetectable expression of the gmdj408b20c gene in all samples on this panel.

EQUIVALENTS

Although particular embodiments have been disclosed herein in detail, this has been done by way of example for purposes of illustration only, and is not intended to be limiting with respect to the scope of the appended claims, which follow. In particular, it is contemplated by the inventors that various substitutions, alterations, and modifications may be made to the invention without departing from the spirit and scope of the invention as defined by the claims. The choice of nucleic acid starting material, clone of interest, or library type is believed to be a

matter of routine for a person of ordinary skill in the art with knowledge of the embodiments described herein. Other aspects, advantages, and modifications considered to be within the scope of the following claims.